The Dissertation Committee for Shaunak Kar Certifies that this is the approved version of the following Dissertation:

Exploring the design principles of orthogonal transcription control systems

Committee:

Andrew Ellington, Supervisor

Bryan Davies

Hal Alper

David Soloveichik

Jeffrey Barrick

Exploring the design principles of orthogonal transcription control systems

by

Shaunak Kar

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2021

Dedication

To my mom, Sharmila and her incomplete dissertation and what it symbolizes.

To my dad, Sujit, for providing unwavering support.

To my brother Shovik, my first science teacher and has been an inspiration ever since I remember.

To the memory of my beloved grandparents Bibha Ray (Didin), Hasi Rani Kar (Thamma), Jyotish Chandra Kar (Dadu),and Kanak Kanti Ray (Dadunda) whose unimaginable sacrifices provided the entire family with the opportunities to realize their hopes and dreams, and I am no exception.

Acknowledgements

This journey has been a long and arduous one. I remember walking in through the doors of MBB, a fresh graduate with a naïve fascination towards engineering genetic circuits. Now, looking back, I feel fortunate to have been able to pursue my interests and learning so much more along the way.

First, I would like to thank my adviser Andy, for providing me with the opportunity and believing (most of the time) that the task at hand wasn't simply an academic exercise but rather an ambitious effort to define a paradigm in the construction of genetic circuits. In his own way of training, he thrust me right into the deep end, and while I struggled to stay afloat for quite a while, and I can safely say now that I am doing just fine. I still remember what he told me on my first day, "I am not your biggest resource, the lab is". Honestly, at that time, I didn't really understand the importance of those words, but now I realize what exactly he meant.

The Ellington lab is truly a unique place, and that is probably an understatement. It's an assortment of some of the most talented, ambitious, and crazy scientists, who in their own right, dream (and often succeed) of changing the world. Such bold ambition fueled to the passion, is infectious and for better or worse it has played an important role in shaping my own scientific endeavors. To Andy's credit, getting such a bunch of folks together, and enabling the vast breadth of projects, is not an easy feat. His enabling has led to the influx

of diverse ideas and skills which have been central to so many great ideas developed by the lab over the years, and my dissertation is no exception.

I am particularly grateful to Elizabeth Gardner a talented scientist, a wonderful colleague and if possible, an even better friend (and definitely not just a rubber duckie!). I fondly look back at all the times we spent discussing broad goals for the 'field' of synthetic biology to behavioral economics, not knowing at the time how these provided the intellectual ammunition for me to plow ahead. Her contributions and insights go far beyond just figures in a paper or assisting with a protocol (she did those things too). So many of the results and discussions laid out in this dissertation, can be traced to the innumerable talks and discussions I have had with her (and continue to do so). Without a formal system for rewarding such efforts, it's rather unfortunate that such contributions can often go unnoticed. Regardless I will always be grateful for her contributions in making this dissertation a reality and in my humble opinion, just authorships on papers doesn't suffice.

To Kamyab, for attempting (and succeeding) to resurrect my rather crazy ideas and making them a reality. I am thankful for your continued support. To Jimmy Gollihar, who arguably is the personification of the lab, for making me feel lazy (unknowingly), and encouraging me to push for even greater heights. To so many past and present members of the lab – Simon, Stella, Raghav, Jared, Ross, Dan, Andre, Dave, Inyup and many more, thanks for making the lab a fun place. To Cody, Michelle and Arti for running the day-to-day operations of the lab, allowing us to succeed in our projects. To Sanchita, for always being a calm presence amidst all the craziness.

To all my friends here in Austin, and around the country – Vivek, Priyanka, Sanket, Anand, Kasturi, Anurag, Shruti, Rajat, Ankit, Fizza, Mandar, Carl, Tushar, Cynthia, Nidhi and many more for making this journey a lot more enjoyable. In particular, thanks to Vivek and Priyanka for making me feel like a part of your growing family, I am grateful for all the support you guys have given me throughout (and at times literally keeping me fed and healthy).

Finally, to all members of my immediate and extended family, who have been an inspiration and provided unwavering support throughout this long journey. This work wouldn't have been possible without them.

Abstract

Exploring the design principles of orthogonal transcription control systems

Shaunak Kar, PhD

The University of Texas at Austin, 2021

Supervisor: Andrew D. Ellington

The last two decades has witnessed an unprecedented growth in our ability to engineer biological systems for a wide range of applications ranging from the development of smart therapeutics, production of valued products and chemicals and engineering crops with programmable traits and much more. At the core of these capabilities has been the design and characterization of synthetic genetic programs that has enabled the predictable programming of cellular behavior and phenotypes. A fundamental challenge in the construction of such circuits and programs is being able to design and model them against a variety of organismal backgrounds, which can be often difficult to predict and can lead to circuit failure when systems are ported across organisms.

Such failure modes can potentially be mitigated by embedding orthogonal modes of transcriptional control and regulation in genetic programs to drive the expression of the circuit components in both prokaryotes as well as eukaryotes. Specifically, in prokaryotes, we demonstrate how an autoregulated network controlling the expression of an orthogonal RNA polymerase – T7 RNA polymerase, can be utilized to precisely express target genes in a highly predictable manner dictated by mutant T7 RNAP promoters. Furthermore, with

the use of a modular architecture we show how such expression systems can be readily ported across diverse prokaryotes. In each species, the relative strength of expression obtained from the T7 RNAP homeostasis circuit is nearly identical, suggesting T7 RNAP driven expression systems can be utilized as predictable cross-species gene expression platform.

In another example, orthogonal transcriptional regulation was engineered in a complex eukaryote (plants) using a programmable transcription factor - dCas9:VP64 and a set of designed synthetic promoters whose activity can precisely regulated with the expression of specific guide RNAs (gRNAs). This strategy was used to construct three mutually orthogonal promoters, allowing multiplexed control of gene expression in plants. Overall, the design strategies and architectures described in this work can be used to explore the design of more complex circuits where the activity of T7 RNAP can be coupled to regulate the activity of dCas9 based transcription to generate circuits operating across kingdoms of life.

Table of Contents

List of Tables xiii

List of Figures xiv

Chapter 1: Introduction 1

Historical roots of synthetic biology 1

Birth of Synthetic Gene Circuits 2

Expansion and Growth of Synthetic Gene Circuits 4

Design principles of Synthetic Genetic Circuits 6

Tools for gene regulation 7

Transcriptional Control 7

Predictability in transcriptionally regulated gene circuits 11

Orthogonal Transcriptional Control and Regulation 13

Chapter 2: In vitro transcription networks based on hairpin promoter switches 18

Abstract 18

Introduction 19

Results 21

Design of hairpin promoter switches: 21

Performance and modeling of hairpin promoter switches: 21

Regulation of hairpin promoter switches: 23

Buffering of inputs for hairpin promoter switches: 25

Construction of larger circuit cascades, a double inverter: 25

Construction of NAND gates: 26

Discussion 30

Materials and Methods 31

Chapter 3: Construction of synthetic T7 RNA polymerase expression systems 45

Abstract 45

Introduction 46

Results 48

Construction of an aTc-inducible T7 RNAP expression system 48

Construction of an inducible and self-limiting T7 RNAP expression system 50

Construction of a T7 RNAP homeostasis circuit 51

Modeling the dynamics of the T7 RNAP expression circuit 54

Discussion 55

Materials and Methods 56

Chapter 4: Design and construction of a modular framework for orthogonal and predictable gene expression in prokaryotes 70

Abstract 70

Introduction 71

Results 73

Design of Universal Plasmid System (UPS) Architecture 73

Characterization of transcriptional and translational elements in UPS 76

Design and characterization of UPS encoded T7 RNAP homeostasis circuits 77

Cross-species activity of T7 RNAP homeostasis circuits 80

Discussion 82

Materials and Methods 83

Chapter 5: Construction of a modular framework for orthogonal gene expression and control in plants 114

Abstract 114

Introduction 115

Results 117

Construction of the APT (Advanced Plant Technology) toolkit 117

Assaying reporter expression using the APT toolkit 119

Development of an Orthogonal Control System (OCS) to regulate transgene expression 120

Inducible gene expression system via the OCS framework 122

Construction of a panel of mutually orthogonal synthetic promoters 123

Construction of complex ratiometric circuits 124

Discussion 125

Materials and Methods 128

Some concluding thoughts 165

References 168

List of Tables

Table 2.1: List of oligonucleotide sequences used 44

Table 4.1: List of genetic parts for the assembly of plasmid backbone 98

Table 4.2: List of elements for the construction UPS genetic circuits 108

Table 5.1: Promoters 145

Table 5.2: Terminators 147

Table 5.3: Coding sequences 147

Table 5.4: gRNA expression cassettes 153

Table 5.5: Plant expression vector backbone elements 156

Table 5.6: List of OCS constructs used 161

Table 5.7: List of Addgene plasmids used in this study 162

List of Figures

Figure 2.1: In vitro transcription from hairpin templates. 33
Figure 2.2: Modeling the in vitro transcription network: 34
Figure 2.3: Regulation of transcription. 35
Figure 2.4: Modularity of hairpin promoter switch 36
Figure 2.5: Buffering of inputs for hairpin promoter switch. 37
Figure 2.6: Construction and analysis of the double inverter circuit. 38
Figure 2.7: Design of NAND gate 39
Figure 2.8: Empirical model to determine the repression efficiency of NAND gate in the presence of varying amounts of inputs. 40
Figure 2.9: Leaky expression of rMG from OFF NAND switch. 41
Figure 2.10: Design of the NAND gate with remote toehold design. 42
Figure 2.11: Performance of the NAND gate with RNA inputs. 43
Figure 3.1: An anhydrotetracycline inducible T7 RNAP expression system. 64
Figure 3.2: Design and characterization of a self-limiting T7 RNAP expression system. 65
Figure 3.3: Homeostatic control over T7 RNAP production. 66
Figure 3.4: Homeostasis circuit utilizing a T7 RNAP variant – CGG-R12-KIRV that recognizes a different T7 RNAP promoter. 67
Figure 3.5: Modeling of the homeostasis circuit variant 68
Figure 4.1: Design of genetic constructs with the UPS architecture 86
Figure 4.2: Assay of a constitutively driven mScarlet-I to compare the variation in copy number. 88
Figure 4.3: Design and characterization of transcriptional control (constitutive promoters) and translational control elements (RBSs) 89

Figure 4.4: Design and characterization of T7 RNAP homeostasis circuit using the UPS architecture..... 90
Figure 4.5: Model describing the dynamics of the T7 RNAP homeostasis circuits. 91
Figure 4.6: Comparison of output expression from T7 RNAP homeostasis circuit to that predicted by a quantitative model. 92
Figure 4.7: Control of output production under T7 RNAP homeostasis circuit using different RBSs driving the reporter expression. 93
Figure 4.8: Characterization of T7 RNAP homeostasis circuits in E.coli MG1655 and Nissle strains. 94
Figure 4.9: Characterization of constitutive promoters in *E.coli* MG1655 and Nissle 95
Figure 4.10 Characterization of constitutive promoters in three species: *Salmonella typhimurium, Serratia marcescens and Pseudomonas putida* 96
Figure 4.11: Characterization of one plasmid T7 RNAP homeostasis circuit in three species -*Salmonella typhimurium, Serratia marcescens* and *Pseudomonas putida.* 97
Figure 5.1: Schematic overview of the APT design-build-test cycle..... 134
Figure 5.2: Schematic demonstrating the assembly of single (A) or multiple (B) transcriptional units into a plant expression vector. 135
Figure 5.3. Characterization of reporter constructs assembled using APT toolkit. 136
Figure 5.4. Characterization of activity of synthetic pATF promoters. 137
Figure 5.5. Evaluation of OCS reporter gene expression in transgenic Arabidopsis plants. 139
Figure 5.6 Design and characterization of gRNA expression modules under the control of Pol II promoters. 140

Figure 5.7. The expression of hdCas9 for both OCS 1-1 and OCS 1-5 was confirmed via Western blot analysis (see Materials and Methods). 141
Figure 5.8. Characterization of an ethylene inducible orthogonal control system..... 142
Figure 5.9. Degree of orthogonality of synthetic promoters. 143
Figure 5.10. Design and characterization of a ratiometric circuit. 144

Chapter 1: Introduction

"What is true for *E. coli* is true for the elephant"

- Jacques Monod

HISTORICAL ROOTS OF SYNTHETIC BIOLOGY

In 1961, following their seminal discoveries about the regulation of the lac operon, Jacques Monod and Francois Jacob posited that the cellular response to fluctuating environmental conditions was primarily governed by networks of regulatory proteins coordinating the expression of genes [1]. In the following decades, as the molecular details of the various regulatory processes were elucidated, it became clear that cells employed a wide array of complex signal transduction and processing pathways to coordinate the expression of genes both at the level of transcription as well as translation [2-4]. The ushering of the 'omics' era in biology catalyzed by high-throughput techniques like microarray technologies, next generation DNA sequencing, provided a more holistic view of gene regulatory networks [5]. It became apparent that the vast number of regulatory proteins and pathways were organized as hierarchy of discrete functional modules, which in turn dictated the gene expression dynamics [4, 5].

In parallel, the advent of the recombinant DNA technology in the 1970s and 80s led to the design and construction of synthetic gene switches [6-8]. These simple transcription switches enabled temporal control of expression by placing the target gene under synthetic promoters engineered to be regulated by naturally occurring transcription factors such as LacI or TetR, and the expression of the target gene can be induced upon the addition of a

specific molecule like IPTG (LacI) or anhydroetracycline (TetR) [9]. The inducers were specifically chosen to act orthogonally to that of the host metabolism, to only selectively express the target gene without perturbing the host chassis. Initially characterized in *Escherichia coli* for recombinant protein expression, such switches were then designed to operate in eukaryotes by simply incorporating these transcriptional regulatory elements in eukaryotic promoters [6, 8]. Such inducible gene expression systems have since proven to be invaluable tools in recombinant protein expression and advanced our mechanistic understanding of biological processes.

BIRTH OF SYNTHETIC GENE CIRCUITS

Moving forward, coupled with the understanding of gene regulatory networks, it posed a tantalizing question – was it possible to design genetic programs beyond just simple transcription switches and embed them in cells to generate predictable phenotypes? That simple question is the central tenet of synthetic biology. The ability to control and alter cellular behavior by embedding them with artificial genetic programs will have profound implications, resulting in the design of smart therapeutics and diagnostics, generation of crops with programmable traits, and biomanufacturing of novel products[10-14]. Shaped by billions of years of evolution, cells have developed a plethora of regulatory processes to modulate and coordinate expression of gene(s) in response to a wide range of chemical and physical stimuli [15]. These regulatory modules can be understood as genetic circuits or programs, and are typically composed of fundamental elements such as promoters, ribosome binding sites (RBSs), terminators, transcription factors, receptors and

so on. The interactions between these elements determine the 'logic' of the genetic program. To better discern the design principles behind these modules and networks, it was critical to construct synthetic versions with programmed interactions and assess their dynamic behavior. Such efforts were further catalyzed by the accelerated growth of recombinant DNA technologies along with plummeting synthesis and sequencing costs allowing for assembling designed genetic constructs.

About two decades ago, two seminal papers were the first to demonstrate how existing transcription factors and their recognized promoters can be rewired to construct artificial network topologies beyond simple gene switches [16, 17]. In both cases these gene networks were designed to execute a specific dynamic process which was predicted by a quantitative model. In one example, Collins and colleagues designed a genetic toggle switch using two repressors – LacI and TetR connected in a mutually inhibitory manner [16]. Using a relatively simple quantitative model describing the mRNA and protein expression dynamics of the circuit elements, the authors showed that cells harboring this synthetic circuit could toggle between two stable states of gene expression upon exposure to a chemical or a physical input. In the other example, a synthetic transcriptional oscillatory network was designed by Elowitz and colleagues using three repressors – LacI, TetR and cI, (lambda repressor) where each genetic element cyclically inhibited the expression from its adjacent promoter [17]. Similar to the characterization of the genetic toggle switch, the authors leveraged the use of a quantitative model to guide the design process and identify the necessary conditions to achieve oscillatory behavior. While both these circuits were built using a handful of genetic elements, these first successful instantiations of synthetic

circuits put forth a new model-guided paradigm for the engineering of synthetic gene regulatory networks in biological systems.

EXPANSION AND GROWTH OF SYNTHETIC GENE CIRCUITS

In the following years, inspired by the toggle switch and oscillator, several studies applied similar circuit engineering principles to build and quantitatively characterize a wide number of network motifs and circuits. Initially, limited by the set of well characterized genetic parts, circuits with relatively simple topologies were characterized. Some notable examples include auto-regulatory motifs such as positive and negative feedback loops, cell-cell communication modules for band-pass filters and pattern-formation [18-21]. As new genetic elements – transcription factors, promoters, RBSs began to be mined, designed and engineered, the size and complexity of the circuits grew significantly [22, 23]. Sophisticated and layered circuits beyond simple gene regulatory modules were demonstrated such as distributed computing modules, genetic counters and Boolean functions [24-26]. In the initial years, owing to our deep understanding and ease of genetic engineering, *E. coli*, served as the ideal chassis for characterization of majority of the circuits. Moving forward, to assess the performance of the synthetic gene networks in different hosts especially in eukaryotes such as *Saccharomyces cerevisiae* and human cell lines, design and characterization of new promoters and transcription factors was required to ensure that circuits operate optimally in the context of the new host chassis [11, 27, 28]. As a notable example, a genetic toggle switch was characterized in mammalian cells implementing the same architecture as the original Collins' design [29]. By using appropriate repressors acting on engineered promoters

characterized in mammalian cells, connected in a mutually inhibitory manner, it was shown that this network can indeed act as a memory element, similar to what was demonstrated in *E.coli* [16, 29]. Thus, it showed that for a given gene circuit, preserving the network architecture encoded in the form of interactions of the genetic elements can yield similar dynamic behavior across diverse hosts. Similarly, a variety of network topologies were transplanted in eukaryotic hosts, such as oscillators, layered digital gates and distributed computing modules [10, 18, 27, 30-32]. In addition to engineering gene circuits to perform complex computations, significant efforts were undertaken to broaden the sensing capabilities of such circuits. In general, allosteric transcription factors that respond to a variety of inputs have been extensively engineered and incorporated as sensors in genetic circuits. Such transcription factors have been shown to operate optimally in both prokaryotes and eukaryotes [30, 33]. Signal transduction pathways in both prokaryotes and eukaryotes were engineered that enabled the sensing of novel inputs [15]. A particularly relevant example was the engineering of a non-native two component sensor to sense light in *E.coli*, which in turn was connected to a genetic circuit to compute the specific functions [34, 35]. In eukaryotes, the plethora of natural receptor mediated pathways such as GPCRs, receptor tyrosine kinases have been primary targets to act as sensors [11, 15]. These have been connected to genetic controllers to execute specific logic operations. More than two decades since the toggle switch and oscillator were published, it is now well established that cells can be engineered to perform complex computations and novel tasks with designer genetic programs.

DESIGN PRINCIPLES OF SYNTHETIC GENETIC CIRCUITS

The successful implementation of a variety of synthetic gene circuits across diverse hosts, prokaryotes and eukaryotes, cemented the notion that circuit engineering principles could be successfully applied for the bottom-up construction of artificial gene regulatory modules with predictable dynamics. Despite the successful demonstration of a number of circuit modules, design of genetic circuits was far from a formal process. Often it relied on an ad-hoc process of identifying the optimal set of genetic parts such as transcription factors and their regulated promoters and further required an iterative process of optimizing the expression and mitigating cross-talk between the various circuit components to produce the desired output in host organisms [36]. To streamline the design process, a gene circuit has been broken down into three main functional modules – sensors, control system and actuators [25, 36]. Briefly, in the overall circuit, sensors transduce the signal (presence of specific chemical or physical input) typically by modulating the activity of a specific promoter, either directly via allosteric transcription factors or through more elaborate receptor mediated signaling pathways. The transduced signals from the sensors then serve as inputs to the control system, which encodes the designed logic function(s) to process the inputs. The control system is encoded as gene regulatory networks, where the logic operations are programmed via the interactions (transcriptional and/or translational) of the various genetic elements. Upon execution of the logic operations, the expression of actuators typically results in the desired cellular phenotype. For purposes of this review, I will focus on the existing architectures for the construction of control systems with a special emphasis on transcriptional regulation.

Tools for gene regulation

The design of artificial control systems in genetic circuits requires the precise regulation of expression of target gene(s). While transcriptional regulation has been the primary mode of control in genetic circuits (explained in detail below), additional modalities of regulation such as post transcriptional and translational control elements have also been used explored to further augment the computing capabilities of the control system. Majority of such strategies have employed RNA based regulators of gene expression that control translation, degradation and processing of target mRNAs [37]. In particular, these regulators rely on specific RNA motifs that can be programmed to interact with target mRNA either in *cis* or *trans*, governed by simple Watson-Crick base pairing. For translational control, designed RNA hairpins block translation by sequestering the RBS of the target gene [38-40]. Then, upon binding of a trans-activating RNA the translation of the gene can be upregulated. Since the activity of riboregulators can be easily programmed using simple Watson-Crick base pairing rules, it has led to development of orthogonal translational control elements, that operate solely on sequence specificity [39]. Remarkably, the dynamic range of expression of these engineered riboregulators was comparable to those observed for traditional protein-based transcription factors.

TRANSCRIPTIONAL CONTROL

One of the most common and widely adopted frameworks for the construction of biological control systems, relies on the programmed transcriptional regulation of specific promoters which serve as the fundamental unit of computation in a genetic circuit. In such a framework, both the input and output of the control system are defined in terms of

transcriptional activity of specific promoters, where the signal is quantified in terms of RNA polymerase flux associated with each promoter. Since direct measurements of flux can be tedious, it is generally inferred by the expression of a reporter protein placed under the control of the promoter [36]. The transcription activity of a given promoter can be regulated via transcription factors by either recruiting the RNAP (activators) or sterically blocking the binding or progression of RNAP (repressors). Naturally occurring transcription factors recognize specific sequences of DNA – termed as operators, and these operators can be modularly combined with host RNAP recognition sequences to design synthetic promoters with predictable modes of regulation [9, 22, 25, 26]. This framework also enables the design of sophisticated programs composed of layered circuits, where the expression of the specific transcriptional regulators can be placed under the control of an engineered promoter of choice and thus, in such a configuration, an engineered promoter can be designed to regulate the activity of another promoter by simply expression the requisite transcription factor. As mentioned previously, in its simplest instantiation, synthetic transcriptional control was utilized to construct gene switches. In theory, the size and complexity of a transcriptionally controlled circuit is limited only by the number transcription factor: promoter pairs that can operate in the target organism with minimal crosstalk. In the initial years, the availability of only a limited set of well characterized transcription factors and inducible systems, relatively simple network and circuits were explored. For the construction of more complex circuits typically consisting of sequential layered regulation, significant efforts have been undertaken to alleviate this major bottleneck. Voigt and colleagues identified a panel of TetR homologs and characterized

their respective promoters, that showed minimal cross-reactivity [22]. The activity of each of the repressor was quantified as a Boolean NOT gate, i.e the addition/expression of the repressor (input) results in the inhibition of transcription from the engineered promoter (output). By identifying a set of mutually orthogonal repressors and their respective promoters in *E. coli*, the authors also demonstrated how such NOT gates can be stitched together to genetically encode for larger Boolean functions [22]. Such hierarchical assembly of complex circuits from simple gates has been one of the hallmarks of synthetic circuit design. It is important to note that in large circuits, each repressor-promoter pair required precise tuning of their expression to obtain the required response. This is typically achieved with the use of mutant promoters (based on host RNAP) as well as translational control elements (mutant RBSs). Beyond mining naturally occurring transcription factors, rational design of DNA binding proteins has proven to be an effective strategy for the de-novo construction of transcription factors. Zinc finger proteins and TALEs served as ideal modular programmable protein motifs, whereby simply changing the amino acid sequence, these proteins could be targeted to bind any DNA sequence of choice[41]. These engineered DNA binding proteins can act as either as repressors, or as activators by fusing modular regulatory domains [41].

In eukaryotes, along with the naturally occurring transcription factors, modular DNA binding domains have been extensively used for the construction of synthetic transcription factors and their corresponding promoters [42]. Unlike prokaryotes, the design of synthetic promoters in eukaryotes follows a more universal architecture where the DNA binding sites is followed by a minimal promoter. The latter acts an efficient initiation signal

for the RNA polymerase II (Pol II) and the DNA binding sites act as landing pads for the artificial transcription factors [42]. The artificial transcription factors are typically composed of a DNA binding protein fused to modular activation domains. Such a formal architecture for the construction of artificial transcription factors and their respective promoters has enabled the bottom-up construction of large complex genetic circuits [42].

More recently, with the discovery of CRISPR, a class of DNA binding proteins with a novel regulatory modality were engineered [43]. The nuclease 'dead' version of the Cas9 (dCas9) was shown to bind to DNA sequence determined by the specificity of small guide RNAs (gRNAs) [43]. By simply designing gRNAs to target specific regions of a promoter, it was shown dCas9 can efficiently repress transcription from the promoter, a phenomenon commonly termed as CRISPR interference (CRISPRi) [43]. Apart from repression, these proteins can also be coupled with modular activation domains such as omega factors (prokaryotes), VP64 or p65 (eukaryotes), to act as transcription activators [44, 45]. Since the specificity of dCas9 based transcription factors can be programmed based on predictable Watson-Crick base pairing rules, it has enabled the construction of layered logic gates, where a panel of promoters were designed with minimal regulatory cross-talk by simply changing the guide RNA sequence [27]. In addition, the ability of dCas9 based transcription factors to alter transcriptional activity of endogenous genes in both prokaryotes and eukaryotes has provided a powerful tool in the creation of transient phenotypes with the use of multiple gRNAs generated as outputs of control systems [44, 46].

Predictability in transcriptionally regulated gene circuits

One of the principal goals of genetic circuit design is to predict the behavior of a designed gene circuit under diverse conditions and in different host chassis. Ideally, the behavior of a given synthetic circuit can be fully derived by modeling the interactions of the various circuit components. Specifically in the case transcription control systems, the use of orthogonal transcription regulators has greatly improved in the predictability of dynamics of a gene circuit. However, since all synthetic gene circuits utilize host resources for its execution such as RNAPs and ribosomes for the transcription and translation respectively, the performance of any circuit is inherently intertwined with the host regulatory processes [36]. It has been shown that even slight changes in growth conditions such as media composition, temperature or differences in the host genetic context (strains) can affect circuit performance [36, 47, 48]. Even in a well characterized host such as *E. coli*, the dynamics of gene expression can be highly variable due to slightly different strain backgrounds. Specifically, in the case of transcriptionally regulated circuits, the dynamic behavior of a designed promoter regulated by specific transcription factor can be mapped to what is described as a response function [36]. The response function provides describes the activity of the promoter (in terms of a reporter gene) in the presence of varying amounts of transcriptional regulator. Thus, the overall transcription response of the circuit behavior can be described as a summation of all the response functions of promoters: regulator pair. While such a simplified abstraction has been useful in predicting circuit dynamics, there are still significant challenges that lay ahead. To improve predictability and consistency in the characterization of gene circuits, a number of strategies have been adopted. First, it was

proposed that the performance of each genetic part was determined relative to a constitutive expression control, and the activity of each element was expressed as relative expression units. Such a strategy was shown to mitigate any global perturbations [49]. This was shown to greatly improve the accuracy of genetic parts even when they were characterized under diverse conditions. Such a framework was adopted by Voigt and colleagues in the construction of an automated framework for the design and characterization of an arbitrary Boolean functions assembled from characterized promoter gates [22, 25]. In addition, it was observed that the compositional context of any genetic element that can arise while assembling of parts into a plasmid backbone or in the genome can lead to undesired interactions and affect circuit performance [36, 50]. Such effects can be minimized by implementing modular cloning strategies to minimize composition context of the assembled genetic circuits, and incorporating insulator sequences such as ribozyme processing elements, strong sequence diverse terminators. An elegant demonstration of these strategies was shown by Voigt and colleagues in creating an automated framework for the design of arbitrary Boolean gates based on characterized response functions of transcription regulator: promoter pairs[25]. Overall, the authors demonstrated the construction a total of 60 Boolean circuits in *E.coli* circuits (consisting upto 10 regulators and 55 genetic parts), and out of which 45 performed optimally where the output can be accurately predicted by simply summing up the response functions of each promoter gate [25]. More recently, the authors have successfully implemented a similar framework for the design and construction of Boolean gates consisting upto 11 different regulators in *Saccharomyces cerevisiae* [30]. Furthermore, they have extended the quantitative model in

predicting the dynamic behavior of these circuits over days [30]. However, most of the strategies described have been effective in predicting the performance of the circuits in only a few organisms (both prokaryotes and eukaryotes), often model organisms. Currently, when such circuits are ported across diverse hosts, it is often to design new control elements, necessary to characterize all circuit in the new host background before a successful prediction can be made. This has posed major limitations for the robust portability of designed synthetic circuit across different host chasses.

ORTHOGONAL TRANSCRIPTIONAL CONTROL AND REGULATION

As stated previously, the predictability of gene circuits and specifically that of control systems, relies on the orthogonal behavior of the circuit components. In the case of transcriptionally regulated circuits, given that the transcription factors and their respective promoters interface primarily with the host RNAP, it can often result in unpredictable interactions that can greatly limit the performance of the circuit when ported across species[36, 51]. This often results in circuit failure when the circuits are transferred across different species. Thus, to ensure robust portability, there is a need for characterization of genetic parts that operate orthogonally to the host RNAP itself. In prokaryotes, this is typically achieved with the use of single subunit phage RNA polymerases, most commonly T7 RNA polymerase (T7 RNAP). Owing to highly robust and orthogonal activity towards it small 17bp promoter, the use of T7 RNAP has typically been used in the over-expression of target genes [52]. By decoupling host transcriptional regulation of the target genes, in theory, arbitrary levels of expression can be achieved simply controlling the levels of T7

RNAP across diverse species [53, 54]. However, due to its expression mediated toxicity, precise control over the expression of T7 RNAP is required for the construction of more complex control systems. Typically, this has been achieved with the use stringent inducible systems in addition to translational control elements to limit the levels of T7 RNAP[55 56]. Such a strategy has been shown to limit the expression of T7 RNAP to subtoxic levels, and has been shown to be effective in establishing orthogonal control of metabolic pathways [56]. Apart from the high expression capacity of T7 RNAP based systems, decades of biochemical and mechanistic understanding of the enzyme has enabled its engineering paving the way for a variety of novel applications [57-60]. Notably, a panel of six highly orthogonal T7 RNAP variants and promoter pairs were engineered using directed evolution were engineered multiplexed control of gene expression [61, 62]. A fragmented design of T7 RNAP was also published and shown to be an effective resource allocator [63]. More recently, split versions of T7 RNAP have also used for the study protein: protein interactions in a manner similar to yeast two hybrid system [64]. Beyond the engineering of T7 RNAP, new modalities of control over the expression of the target gene(s) can be introduced, by engineering existing transcriptional repressors to repress the T7 RNAP promoter, thus establishing an orthogonal sub-system of a transcriptional regulatory system operating independently within a host. Such frameworks powered by T7 RNAP as the transcriptional engine will therefore mitigate the crosstalk between the synthetic circuit components and the host RNAP, thereby robustness of synthetic circuits.

However, since in most cases, the expression of T7 RNAP itself is placed under host transcriptional regulation, it has limited the predictability and portability of such

systems across diverse prokaryotes. To this end, Salis and colleagues proposed a novel autoregulatory module, where the expression (positive feedback) and subsequent regulation (negative feedback via TetR) of T7 RNAP is placed under its own control, thereby insulating it against any host RNAP dependent regulation [65]. Furthermore, the authors showed that by simply adjusting the strength of the feedback loops, different steady-state levels of T7 RNAP can be achieved, thus modulating the 'orthogonal transcriptional capacity' of such a system [65]. Moving forward, these so called T7 RNAP homeostasis circuits provide an attractive platform for the bottom-up construction of truly orthogonal transcriptional control systems in prokaryotes. Furthermore, the expression of any target gene(s) under the control of T7 RNAP can be predictably modulated with the use of mutant promoters. Owing to these orthogonal interactions, such transcriptional regulation can be effectively ported across diverse species without any loss in predictability. However, in eukaryotic hosts, while T7 RNAP has been shown to be functional, the lack of post-transcriptional processing of the generated transcripts leads to poor translation of these mRNAs [66]. This lack of translation of T7 RNAP derived transcripts has limited the use of T7 RNAP as a transcriptional engine.

However, unlike in prokaryotes, the design of synthetic promoters in eukaryotic utilizes a common architecture, where a strong transcriptional initiation region is cloned downstream of DNA binding operator sequences [42]. The latter act as landing pads, and can be readily programmed to be orthogonal to the host chassis [42]. This minimizes any unwanted cross-talk with endogenous regulatory elements and ensures the insulation of the synthetic promoters. Owing to availability of universal modular activation and repressor

domains, artificial transcription factors can be designed to possess predictable modes of regulation. Remarkably, it has been shown that activation domains such as domains (VP64, p65, Rta) when used in tandem, possess strong activation across diverse eukaryotes[67]. Thus, by combining such 'universal' activation domains, with DNA binding proteins to generate artificial transcription factors (ATFs) that can operate across diverse eukaryotes. These ATFs can be designed to be orthogonal to the host regulatory by selecting the operator sequences to be divergent from the host genome and in parallel the corresponding synthetic promoters can be designed such that each promoter is regulated by an orthogonal artificial transcription factor. Naturally occurring DNA binding proteins sourced from prokaryotes such as LacI and TetRs have been previously shown to be orthogonal to the host [6, 68]. More recently, various protein templates have been utilized for the de-novo construction of transcription factor with user defined operator sequences. These include zinc finger proteins, TALEs and CRISPR associated proteins (dCas9)[42, 69, 70]. The latter provides an attractive platform, where a single artificial transcription factor can be utilized precisely regulate multiple promoters simultaneously by simply controlling the expression of specific gRNAs [67, 70]. In this case, the synthetic promoters controlled by dCas9 based transcription factors where the promoters are designed to contain orthogonal gRNA binding sites [71].

In particular, the ability to regulate the activity of these transcription factors (and thereby the activity of promoters) via the expression of gRNAs has also enabled the use of T7 RNAP to express the specific gRNAs to regulate the activity dCas9 based transcription factors [72]. Taken together, with the programmability and near universal activity of both T7

RNAP and dCas9 based transcription factors, a hybrid orthogonal transcriptional control system operating independent of the host transcriptional regulation in both prokaryotes and eukaryotes can be envisioned.

Chapter 2: In vitro transcription networks based on hairpin promoter switches

In this chapter, I begin by characterizing an *in vitro* transcription network, powered by T7 RNAP, where the regulation of each promoter is programmed by the predictable strand exchange logic governed by simple Watson-Crick base pairing.

This chapter is reproduced (with minor modifications) from its initial publication:

Kar, Shaunak and Ellington AD (2018) In vitro transcription networks based on hairpin promoter switches *ACS Synthetic biology*.

Author contributions:

S.K and A.D.E conceived of the idea. S.K designed, performed the experiments and analyzed the results with input from A.D.E. Both authors wrote the manuscript.

ABSTRACT

In vitro transcription networks are analogs of naturally occurring gene regulatory networks that consist of synthetic DNA templates that are cross-regulated by their own transcripts. This ability to design and execute *in vitro* transcription networks has allowed bottom-up construction of complex network topologies with predictable dynamic behavior. Here we describe the simplified design of an *in vitro* transcription network based on single-stranded synthetic DNA hairpin switches that function similar to molecular beacons, via toehold mediated strand displacement. Systematic construction of increasingly larger circuits was achieved by programming interactions between multiple switches through rational sequence design, and the

dynamic behavior of networks was accurately predicted using a simple mathematical model. Ultimately, we engineered a cascade of switches that acted as a Boolean complete NAND gate capable of sensing both DNA and RNA inputs. The tools and framework that have been developed makes the execution of *in vitro* transcription circuits much simpler, which will enable them to more readily serve as testbeds for nucleic acid computations both *in vitro* and *in vivo*.

INTRODUCTION

The field of synthetic biology aims to develop principles for biological circuit design, in part through the use of interchangeable parts to optimize transcription and translation of target genes or pathways [36, 73]. This has often been achieved by using inducible transcription factors along with promoter engineering [25], and has led to the construction of increasingly complex information processing circuits *in vivo*: bi-stable switches [16], oscillators [17, 19, 74], pattern generators [21], logic gates [25, 75, 76] and analog circuits [77, 78]. Circuits have proven useful both in cells, and in acellular reaction networks [79, 80]. However, the lack of predictability of the performance of parts, and their potential toxicities can lead to failures in circuit performance [36, 50, 81].

In parallel, the emergent field of DNA computing has led to the *de novo* construction of a variety of molecular machines and logic circuits [82-87] that rely primarily on simple and predictable Watson-Crick base-pairing rules. In particular, these purely nucleic acid based systems often use toehold-mediated strand displacement to implement programmable chemical reaction networks capable of performing complex information processing [82, 87-90]. Such networks can be coupled with purified enzymes to catalyze the dynamic production and degradation of

nucleic acid signals, and hybrid nucleic acid:enzyme computing systems have been used to construct systems that mimic naturally occurring gene regulatory networks such as bi-stable switches [91-93], oscillators [94, 95], molecular clocks [96], and circuits capable of homeostasis [97].

In vitro transcription has proven to be an especially convenient platform for the construction of circuits and networks [91, 92, 95-98]. In this framework, the transcription of synthetic DNA templates is regulated by transcribed RNA signals [91, 95]. Because of the predictability of base-pairing and the stability of nucleic acid secondary structures, performance can be modeled [99, 100] and finely tuned over orders of magnitude [88, 90, 101]. Transcription networks therefore provide a direct link between modular parts and the ability to compute [91, 95, 102].

Unfortunately, the modular parts in this framework are predominantly multi-stranded complexes, and this poses a challenge in scaling such systems, especially since each gate must be constructed and purified separately, a feat which stands little chance of being conveniently adapted to cellular regulation [87]. To address this, we have greatly simplified the design of *in vitro* transcription switches by constructing and characterizing a set of synthetic hemi-duplex hairpin transcription switches [102]. Each switch is typically 'ON' in its ground state, producing transcript, but can be inactivated by RNA-binding in a fashion similar to a molecular beacon. Because of the simplicity of design, we were able to readily construct numerous derivatives, including stacked switches and gate structures, and to readout performance by following the production of a fluorescing RNA, malachite green aptamer. As reaction parameters for the basic switches were obtained, the performance of ganged switches and circuits proved to be highly predictable.

RESULTS

Design of hairpin promoter switches:

In order to simplify the design and ultimately the construction of nucleic acid switches for computation, we have developed single-stranded hairpin structures that can be transcribed by T7 RNA polymerase and whose activity can be modulated in a molecular beacon-like way (**Figure 2.1**). These hairpin promoter switches consist of three major domains: a double-stranded (ds) T7 RNAP promoter, a regulatory domain encoded in the 18 nt loop of the hairpin, and a mostly single-stranded (ss) template strand encoding the output of the switch. The modularity of the domains allows us to readily program interactions between different switches based solely on simple Watson-Crick base pairing.

The molecular-beacon like change in state programs the switch. In the ON state, the switch has a fully double-stranded T7 RNAP promoter, which in the presence of the polymerase leads to efficient expression of the downstream output RNA signal. The loop of the hairpin serves as a 'toehold', a single-stranded DNA domain that a complementary oligonucleotide can bind to and in turn open up the hairpin, much like a molecular beacon. This disrupts the T7 RNAP promoter, making it partially single-stranded, and puts the switch in the OFF state.

Performance and modeling of hairpin promoter switches:

In order to gauge the utility of our design, we analyzed the kinetics of transcription from the ON state of a hairpin promoter switch (H1_tr). The hairpin switch was designed using the 18nt loop sequence and T7 promoter sequence (23nt) that were originally proposed that described the theoretical utility of hairpin switch structures for molecular computation. Transcription was monitored by expressing an aptamer (rMG) against the chromophore

malachite green (MG). When MG is bound to the aptamer it 'lights up' and becomes highly fluorescent, thus allowing real time quantitation of transcriptional activity [96, 103]. These substrates allowed us for the first time to quantify in real-time the kinetics of RNA transcription from a hemi-duplex template.

The combination of H1_tr and T7 RNAP led to sustained expression of rMG without any observable background or leaky expression (**Fig 2.1**). Increasing the concentration of H1_tr for a fixed amount of T7 RNAP (~200nM) led to higher expression of rMG and eventually saturated. The kinetics of the circuit nicely mimic Michaelis-Menten kinetics, in which after substrate saturation of the enzyme there is no further increase in product formation.

We therefore constructed a simple mathematical model to describe the behavior of the *in vitro* hairpin transcription switches using a system of ODEs describing the T7 RNAP-mediated transcription and toehold-mediated strand displacement by specific RNA signals. Relative transcription rates were governed by available T7 RNAP, which in turn was governed by the Michaelis-Menten constants of the substrates (K_m and k_{cat}, **Fig 2.2**). The respective Michaelis-Menten constants were determined using non-linear regression in MATLAB (using the in-built nlinfit routine) from the curve between different initial rates and varying substrate concentrations (**Fig 2.3**). Since the promoter and initiation region for T7 RNAP are the same for all the switches we eventually assume that the Michaelis-Menten constants are same for all hairpin switches.

Moving forward, we chose a non-saturating amount of H1_tr (200nM) for additional studies to ensure that we obtained consistent levels of rMG production even after the addition of more substrate for T7 RNAP (more hairpin switches, as described below). As a measure of circuit performance, we looked at the normalized inhibition of rMG production; the amount of transcript

expressed from H1_tr for a particular circuit after 3 hours of transcription relative to the amount of rMG expressed from H1_tr without any regulation (**Fig 2.3**).

Regulation of hairpin promoter switches:

Regulation of H1_tr can be mediated by binding of a single-stranded oligonucleotide to the loop of H1_tr, leading to disruption of the hairpin and of the ON state. Because of the modular design of the switches, the inhibitor strand can in turn be expressed from another switch (H2_tr) forming a two-node repressor network (NOT gate, **Fig 2.3**). Specifically, in this two-node repressor network the output RNA signal of H2_tr encodes for the complementary strand for the loop of H1_tr and is also partially complementary to the T7 RNAP promoter (**Fig 2.3**). Analysis of the respective secondary structures of both the hairpin switches using the NUPACK suite of algorithms showed no potential interactions between the switches at 37°C.

The addition of H2_tr, even at sub-stoichiometric amounts relative to H1_tr, leads to repression of rMG production (Fig 2C). When stoichiometric amounts of H2_tr (200nM) were added relative to H1_tr, around 90% repression of rMG expression was obtained after 3 hours of transcription. In contrast a switch expressing a non-specific RNA signal (H2_Scr) at the same concentration did not inhibit transcription from H1_tr (**Fig 2.3**).

Toehold-mediated strand displacement can be modeled as a bimolecular reaction in which toehold binding is the rate-limiting step and the overall rate constant is governed primarily by association with the toehold sequence. The rate constants for toehold-mediated strand displacement were estimated from one of the repression curves for H1_tr (200nM) + H2_tr (100nM) (**Fig 2.3**). Since the rate of toehold-mediated strand displacement depends on the length

and sequence of the toehold, we again assumed that this rate was the same for all the switches (which have the same length and similar GC content).

Further, utilizing the modular architecture of these switches we constructed the same two-node repressor network by swapping the loop and output domains of the switches (H2_tr_MG and H1_H2_tr) (**Fig 2.4**). All loop domains of the switches were designed to contain similar GC content, to keep the rates of strand displacement uniform for all the switches. The transcription was assayed using the same rMG reporter, and we see that we are able to obtain similar repression curve (**Fig 2.4**).

We also observed some 'leaky' expression of rMG that increased linearly with time, with a rate independent of H2_tr concentration (**Fig 2.3, 2.4**). This was not surprising, as leaky expression has been a constant problem in nucleic acid circuitry that affects performance. For example, in the case of an optimized, non-enzymatic 'see-saw' OR gate operated at 30°C, the amount of leak reaction observed was 20% after 15 hours [84].

In our case, we speculate this leaky expression probably arises due to unstable transient interactions between the partially single-stranded T7 RNAP promoter domains of two switches that are otherwise in the OFF configuration. A model of the leak as a bimolecular reaction estimated that background transcription should be around 9% of the ON signal after 3 hours, in agreement with what was previously reported for the inhibition of the same hairpin switch expressing a different transcript (8.5% in 2 hours) [102]. The modeled leak was taken into account in extracting parameters describing the dynamics of this circuit (**Fig 2.3, 2.2**).

Buffering of inputs for hairpin promoter switches:

The ability to bind and inhibit regulatory sequences is a key function in many genetic circuits. Competitive binding between regulatory sequences, their targets, and an antisense 'sink' has been utilized in DNA strand displacement-based systems to mitigate leaky expression where the sink consumes some portion of the regulatory sequence [83, 84, 91, 95]. For *in vitro* transcription switches it was shown that competitive binding can be used to create ultrasensitive responsivity. Such sinks are essential for constructing circuits with complex dynamical behaviors [91, 95].

We therefore developed a circuit in which repression of a signal was governed in part by co-expression of an anti-sense to rInh1. In such a circuit, rInh has two targets for binding: H1_tr (inhibition) and rInh1_anti (**Fig 2.5**). The amount of rInh_anti acts as a competitor and increases the concentration of RNA that needs to be present or expressed to inhibit the target switch (H1_tr). To integrate rInh_anti into the transcription circuit, we developed an additional hairpin switch structure (H2_tr_anti) that would produce this antisense RNA. To avoid interaction between the single strand regions of switch H1_tr and switch H2_tr_anti (which of necessity must share complementary sequences) we made H2_tr_anti a completely double-stranded template. When rMG production dynamics were examined for this circuit as a function of increasing amounts of anti-sense switch (H2_tr_anti) added, the amount of repression of rMG production monotonically decreased, as expected (**Fig 2.5**).

Construction of larger circuit cascades, a double inverter:

As is the case with other nucleic acid circuitry, the simplicity and specificity of Watson-Crick base-pairing enables the facile construction of larger transcriptional circuits by programming specific interactions between different switches. Previously, we saw that the two-

node repressor network (described above) acts as a NOT gate, thus we hypothesized that a simple cascade in which two such NOT gates were connected would act as a double inverter circuit, leading to activation rather than repression of signal production.

We designed a three-node, double inverter circuit consisting of three serially connected hairpin promoter switches: H1_tr, H2_tr, and H3_tr (**Fig 2.6**). Again, all loop domains of the switches were designed to contain similar GC content. In this cascade the transcription repression of the H1_tr by H2_tr should be limited because of simultaneous inhibition of H2_tr by H3_tr, and we indeed see that with the addition of stoichiometric amount of H3_tr (compared to H1_tr) a decrease in rMG production of only 24% was observed compared to about 85% without the addition of H3_tr (**Fig 2.6**). Some repression is still observed because of initial transcription from H2_tr, which leads to essentially irreversible repression of transcription from H1_tr.

By using the rate constants estimated from the two-node repressor network, the dynamics of rMG production could be simulated. The predicted relative repression in rMG production was 22% in good agreement with our experimental observation of 24% (**Fig 2.6**).

Construction of NAND gates:

While complex analog circuits such as bi-stable switches and oscillators have been built solely using NOT gates, the construction of more logically complex circuitry required a switch that could process multiple inputs. To this end, we attempted to create a Boolean complete NAND gate. We hypothesized that the single inhibitor strand could be split into two separate inputs, each of which would target half of the toehold and part of the T7 RNAP promoter. Cooperative binding of both inhibitor strands should lead to disruption of the hairpin similar to

the NOT gate described above (**Fig 2.7**). For each input (inhibitor) strand the length of the toehold was fixed (9nt) and the length of the regions targeting the T7 promoter region (11nt) were chosen such that the OFF state should be thermodynamically favorable in the presence of the input / inhibitor (as determined by NUPACK).

To assay repression efficiency, increasing amounts of inputs (I1 and I2) were added to the hairpin switch H1_tr (50nM) and the relative rMG expression was determined. Surprisingly, even in the presence of excess inputs strands (200nM) only 59% repression was observed (**Fig 2.7**). Pre-annealing with I1 and I2 prior to initiating transcription did not greatly improve performance (65% repression).

Although these results highlight the difficulty in using thermodynamic considerations for the design of transcriptional switches, since the role of the polymerase itself in mediating promoter formation cannot be readily included, we nonetheless attempted to see to what extent a simple model that incorporated the leak reaction could capture the degree of repression efficiency (**Fig 2.8**). For the model, we assume that at any time T, the amount of rMG produced is proportional to a combination of transcription rates for the ON state and the background observed for the OFF state (**Fig 2.8**). The transcriptional rate for the OFF state alone was calculated by measuring rMG expression following annealing of the NAND gate with excess inputs. Furthermore, since repression efficiency did not improve even with prior annealing of the inputs with the gate, we assume that the kinetics of binding and conformational transitions are relatively fast and that the system reaches equilibrium quickly. Thus the relative levels of ON and OFF states at any time t can be estimated using thermodynamic modeling via NUPACK.

The modeled relative rMG levels obtained for the NAND gate for varying amounts of inputs added proved to be within 2% of what was observed experimentally (**Fig 2.7**).

We hypothesized that at least a portion of the leak from the OFF state might be due to interactions between the two OFF states that led to the formation of an active promoter (**Fig 2.9**). The two input strands targeted complementary domains of the T7 RNAP promoter region (11nt). Upon binding, this would have revealed a significant region (12nt) of the promoter and its complement as single strands, thus potentially allowing interactions between two OFF states to form an active promoter (**Fig 2.7, 2.9**). To mitigate this potential interaction we designed a new input 1 strand (I1.N) that targeted a different region of the T7 RNAP promoter compared to I1 (**Fig 2.10**). The length of the promoter targeted by I1.N was 16nt, and included the 12nt that was previously hypothesized to participate in non-specific interactions in the OFF state (**Fig 2.10**). The length of the new input 1 (I1.N) was again chosen such that the OFF state should be thermodynamically stable in the presence of both input strands (I1.N and I2). We also included a spacer sequence of 6nt that separated the toehold and the promoter region targeted by I1.N, further increasing the stability of hybridization and thus of the OFF state (**Fig 2.10**). This design should limit non-specific activation due to fraying, and was similar to so-called remote toehold designs for nucleic acid circuits in which toehold and branch migration domains are separated by a spacer sequence [104]. However, the presence of the spacer was been shown to slow the strand exchange reaction [104], which could also prove problematic for the execution of our transcription gates.

We first analyzed the performance of the system using DNA inputs. The transcription gate (50nM) was pre-annealed with different concentrations of inputs and rMG expression was

assayed. In the presence of excess (200nM and 400nM) of both input strands efficient repression (97%) was observed, while no repression was observed with a single input alone (**Fig 2.10**). More significantly, there was a 10-fold reduction in leak compared to the previous design (~3% compared to 35% observed for the previous design, **Fig 2.10**), thus validating our initial hypothesis that the leak was indeed due to interactions between the OFF states via partially single-stranded regions of the T7 RNAP promoter. To determine if the new remote toehold design executed more slowly, we carried out the same assay without any prior annealing of the inputs. The repression efficiency was again high (80% and 90% respectively for 200nM and 400nM inputs), albeit lower than when pre-annealing was carried out (**Fig 2.10**). Thus, we see that the performance of the NAND gate with the new design was significantly better than previously, either when both inputs were pre-annealed (97% compared to 65%) or simply added to the reaction (80% compared to 59%).

Finally, we assayed rMG repression after the addition of the hairpin templates (H2_rI1.N and H2_rI2) that express RNA versions of the input strands (rI1.N and rI2) (**Fig 2.11**). The inhibition assay was done at two different NAND gate (H1_tr) concentrations (50nM and 100nM), and each input template (H2_rI1.N and H2_rI2) was added in a stoichiometric amount relative to the H1_tr gate (50nM and 100nM of both inputs) (**Fig 2.11**). Specific repression of around 80% was observed, with no gradual leak even for extended periods of time (8hrs) (**Fig 2.11**), consistent with our observations for the DNA inputs.

DISCUSSION

Nucleic acid effectors can be readily programmed to assist with circuit execution in cellular systems. For example, toehold riboregulators [38, 39, 105] and guide RNAs for ribonuclear proteins – dCas9 [43, 46] have been used to program complex cellular behaviors. However, the operation of synthetic circuits in cells can conflict with physiological demands for growth, which in turn can lead to lack of predictability and failures in circuit performance [36, 81]. Acellular systems provide an attractive alternative for implementing synthetic circuits, especially given the increasing sophistication and utility of cell-free transcription and translation systems [79, 80, 106]. As examples, cell-free systems have been developed as diagnostics [38, 107], for the on demand synthesis of biomolecules [108-112], and for the industrial scale production of therapeutic proteins. As was the case with cellular systems, nucleic acid effectors have been shown to participate in toehold-mediated strand displacement reactions, coupling to downstream translation and yielding a circuit that senses particular nucleic acid sequences and returns fluorescence [38, 39, 105]. However, these systems so far do not couple via transcription, only translation.

The ability to carry out molecular computations that directly impact transcription greatly enhances the programmability of cell-free circuits, and potentially enables a new suite of diagnostic tools and logical control elements for gene expression. In this regard, the hairpin switches that we have developed are amongst the simplest possible 'logical primitives' that could be used for such circuitry, especially since they do not require the formation of hemiduplexes from multiple strands, as has been the case for many other *in vitro* nucleic acid computations. The base set of computations demonstrated herein (NOT and NAND gates) and the fact that circuit performance can be readily modeled provides a basis for further coupling to cell-free gene

regulation. Most importantly, though, the simplicity of construction provides a means of quickly creating much larger and more complex networks of transcriptional circuits than has so far been possible by relying solely on translation.

MATERIALS AND METHODS

Oligonucleotides: For the hairpin templates, the respective oligonucleotides were ordered as PAGE purified Ultramers from IDT (listed in **Table 2.1**). Oligonucleotides were resuspended at 100 μM concentration in deionized water and stored at −20°C. The concentrations of the DNA were measured by ultraviolet spectrophotometry using the NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE, USA). All transcription templates were built using DNA oligonucleotides obtained from IDT. For making the transcription templates the respective oligonucleotides (25uM) mixed in TAE/Mg .The oligonucleotides further underwent denaturation for 5 min at 95°C before being annealed via slow cooling (0.1°C/s) to 25°C. Annealed oligonucleotides were used directly for *in vitro* transcription reactions and excess was stored at −20°C.

Purification of WT T7 RNAP: WT T7 RNAP were purified by standard Ni-NTA 6×His (N-terminal) methods. The plasmid pQE-WT T7 (containing an IPTG inducible T7 RNAP gene) was transformed in BL21-gold (Agilent). Cells were grown in 2×YT media at 37 °C until reaching OD_{600} ~ 0.7–0.8, at which point 1 mM IPTG was added. Cells were grown 4 h at 37 °C. Following induction, cells were harvested by centrifugation and resuspended in binding buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 5 mM imidazole). Resuspended cells were lysed via sonication on ice using 40% probe amplitude for 2 min (1s ON, 1s OFF). Cell debris was

pelleted by centrifugation (30 min:20,000*g*). His-tagged T7 RNAP was purified by immobilized metal affinity chromatography (IMAC). The lysate was run over 1 ml (bead volume) Ni-NTA gravity column pre-equilibrated with binding buffer. The column was washed with 10× column volumes of wash buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 20 mM imidazole). T7 RNAP was eluted off the column by the addition of 4× column volumes of elution buffer (50 mM Tris-HCl, pH 8.0, 0.5 M NaCl, 250 mM imidazole). Dialysis was performed in final storage buffer (50 mM Tris-HCl, pH 8.0, 100 mM NaCl, 1 mM DDT, 1 mM EDTA). Dialates were adjusted to 1 mg/ml and added to an equal volume of glycerol (final concentration 0.5 mg/ml).

***In vitro* transcription assays**: Transcription reactions contained 40 mM Tris-HCl pH 7.0, 30 mM $MgCl_2$, 6 mM spermadine, 6 mM each NTP, 10 mM DTT, 0.2 μM T7 RNAP, with the templates added accordingly and 25uM malachite green solution. For assay 25uL reaction was added to a 384 well plate (NUNC, Thermo) and malachite green fluorescence (excitation/emission 630/655nm) reading was taken every minute in a Safire monochromator (Tecan).

Acknowledgements:

We thank the current and past members of the Ellington Lab for fruitful discussions. We specially thank Sanchita Bhadra and Cheulhee Jeung for comments on in vitro transcription and Adam Meyer providing constructs and suggestions regarding the expression of T7 RNA polymerase

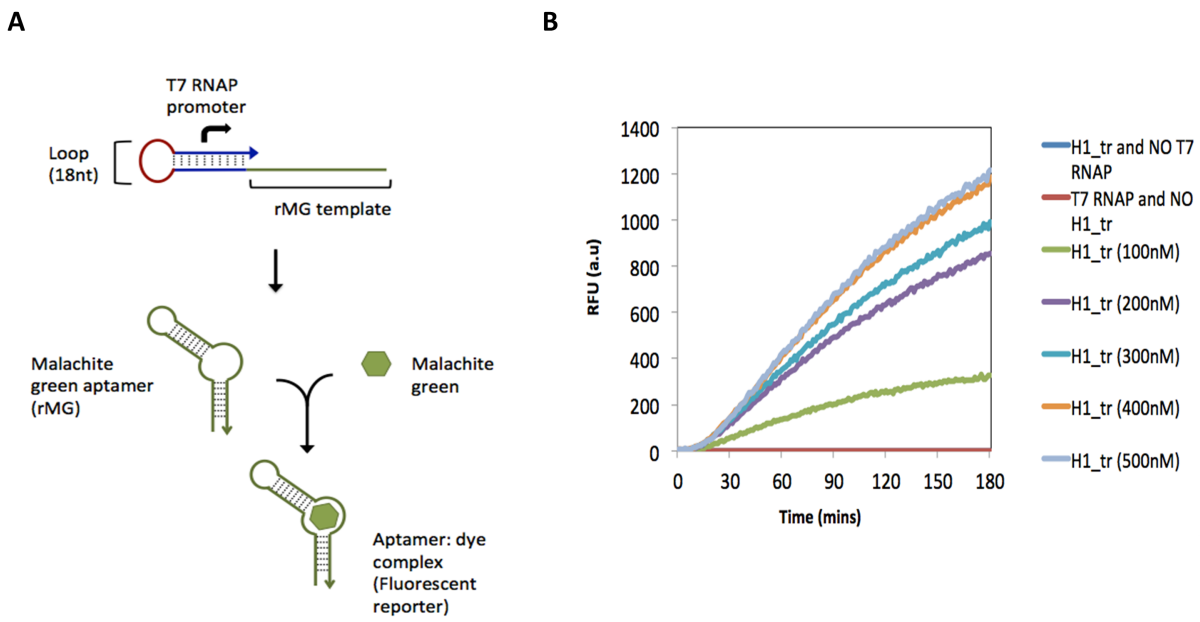

Figure 2.1: In vitro transcription from hairpin templates.

A) Design of the hemi-duplex hairpin transcription template consisting of a 18nt loop, double stranded T7 RNAP promoter and a single stranded template for malachite green aptamer (rMG). B) Fluorescence assay with varying amounts of hairpin template (H1_tr).

Transcription:

$$\text{H1\_tr} + \text{T7 RNAP} \xrightarrow{Km,\ k_{cat}} \text{rMG} + \text{H1\_tr} + \text{T7 RNAP} \quad \text{(a)}$$

$$\text{H2\_tr} + \text{T7 RNAP} \xrightarrow{Km,\ k_{cat}} \text{rInh} + \text{H2\_tr} + \text{T7 RNAP} \quad \text{(b)}$$

Repression:

$$\text{H1\_tr} + \text{rInh} \xrightarrow{kt} \text{H1\_tr\_OFF} \quad \text{(c)}$$

Leak reaction:

$$\text{H1\_tr\_OFF} + \text{H1\_tr\_OFF} \xrightarrow{kleak} \text{rMG} + 2\ \text{H1\_tr\_OFF} + \text{T7 RNAP} \quad \text{(d)}$$

Km, kcat : Michaelis-Menten constants for T7 RNAP

kt : rate of toehold mediated strand displacement

kleak : rate of leaky transcription from OFF switch

Vm : kcat * Et

Differential equations describing the kinetics of the above reactions

$$d[rMG]/dt = (Vm * H1\_tr)/(Km * (1 + H1\_tr + H2\_tr) + kleak * [H1\_tr\_OFF]^2$$

$$d[rInh]/dt = (Vm * H2\_tr)/(Km * (1 + H1\_tr + H2\_tr)$$

$$d[H1\_tr]/dt = -kt * [rInh] * [H1\_tr]$$

$$d[H1\_tr\_OFF]/dt = kt * [rInh] * [H1\_tr]$$

Constant	Value
Km	297 nM
kcat	0.003 /s
Et	200 nM
kt	$3.4 * 10^{-4}$ /nM/s
Kleak	$5.5 * 10^{-7}$ /nM/s

Figure 2.2: Modeling the in vitro transcription network:

Transcription from hairpin templates was modeled using simple Michaelis-Menten kinetics, while the binding of the inhibitor strand leading to toehold mediated strand displacement was modeled as bi-molecular reaction.

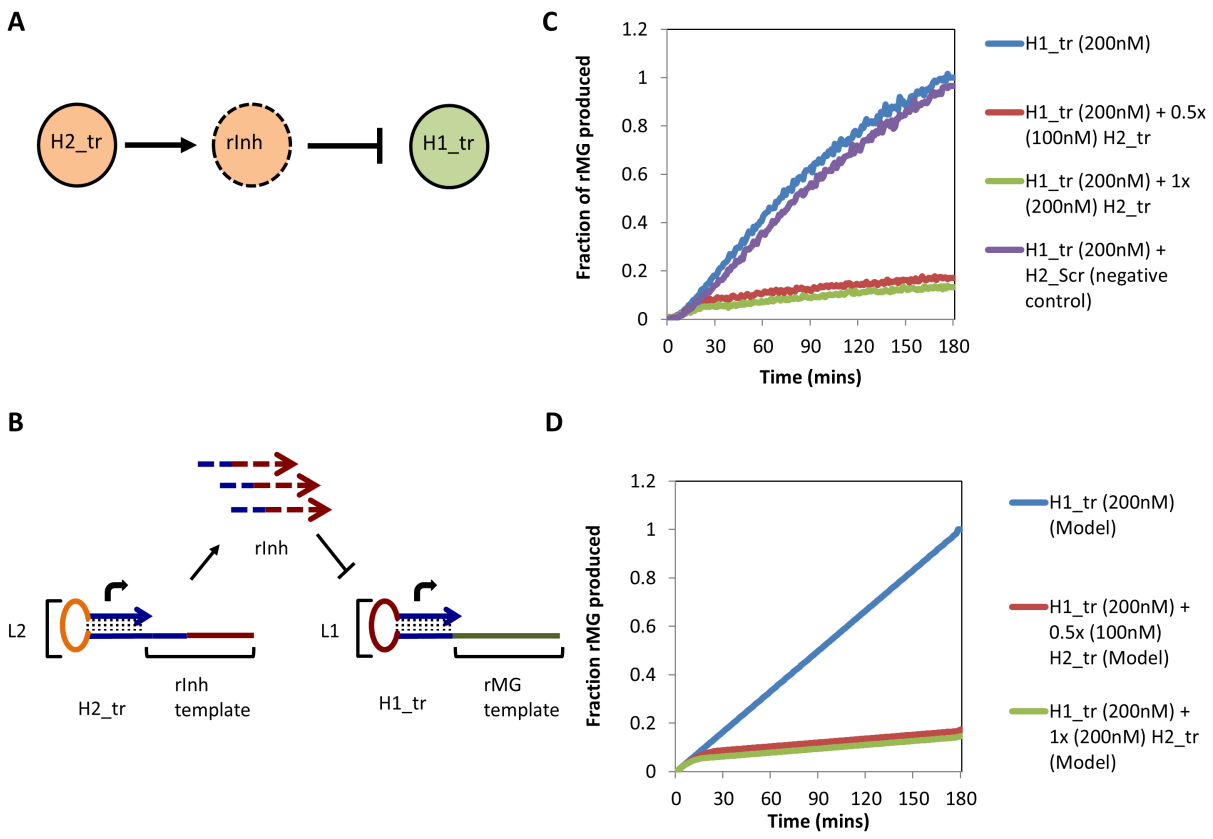

Figure 2.3: Regulation of transcription.

A) Nodal representation of the NOT gate where the transcription from H1_tr is inhibited by the output of H2_tr (rInh). B) NOT gate as implemented in the hairpin switch framework. C) Assay to measure the production of rMG in the presence of different amounts of H2_tr (inhibitor template). D) Simulation dynamics of the modeled reactions involved in the circuit.

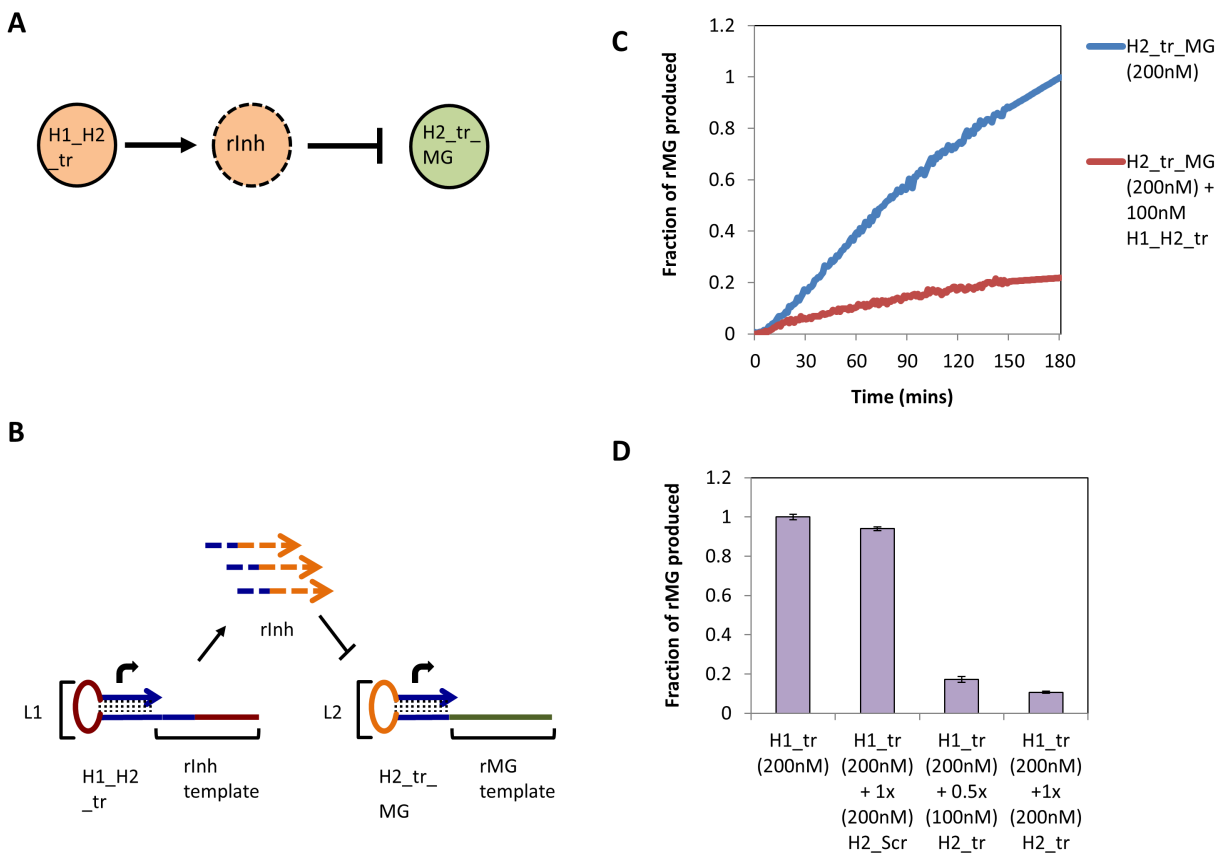

Figure 2.4: Modularity of hairpin promoter switch

A) Nodal representation of the two node repressor network with loops of the hairpin promoter switch swapped. B) The new two-node network implemented in the hairpin promoter switch framework. C) rMG expression from H2_tr_MG in the presence of the H1_H2_tr D) Fraction of rMG expressed after 3 hours for the output of H2_tr repressing H1_tr.

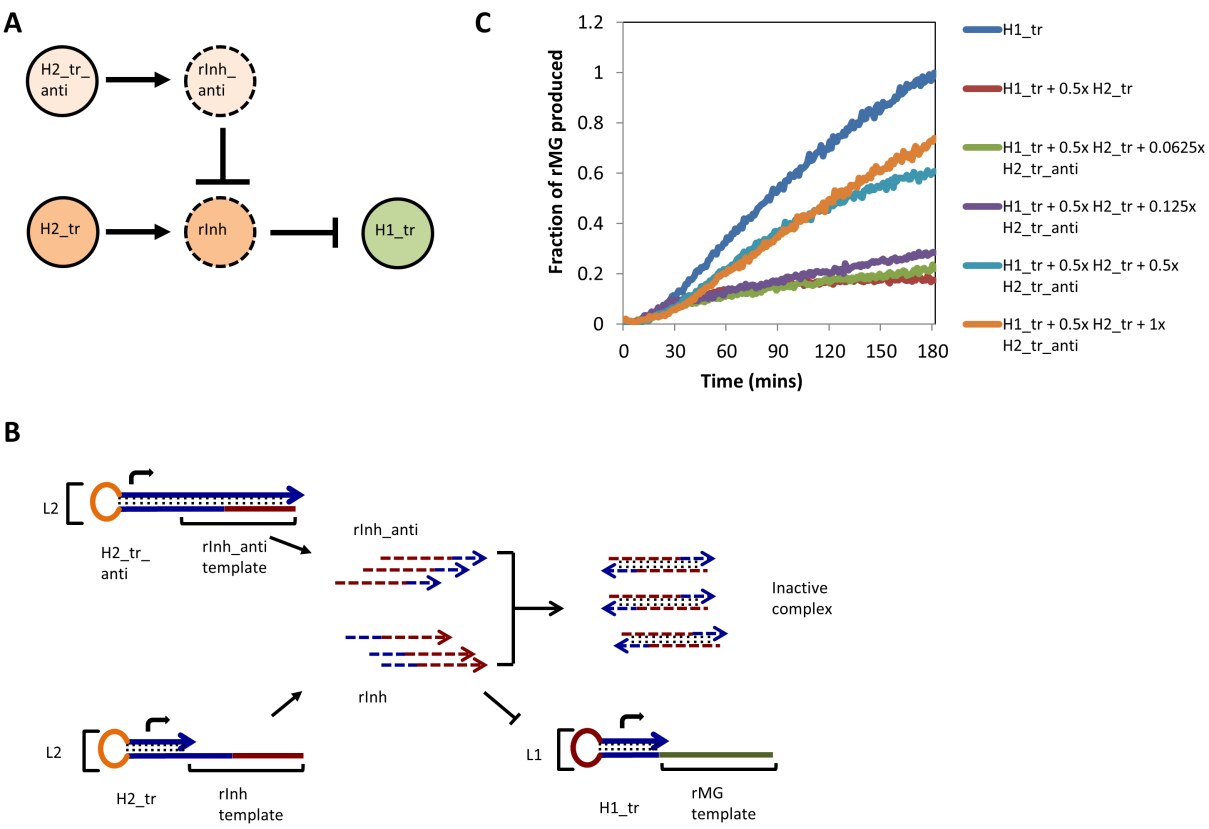

Figure 2.5: Buffering of inputs for hairpin promoter switch.

A) Nodal representation of the buffering network, where the H2_tr_anti expresses the rInh_anti which after binding to rInh renders it inactive. B) The buffering circuit implemented in the hairpin promoter switch framework. C) rMG expression assay from H1_tr (200nM =1x) in the presence of H2_tr (100nM) with varying amounts of H2_tr_anti.

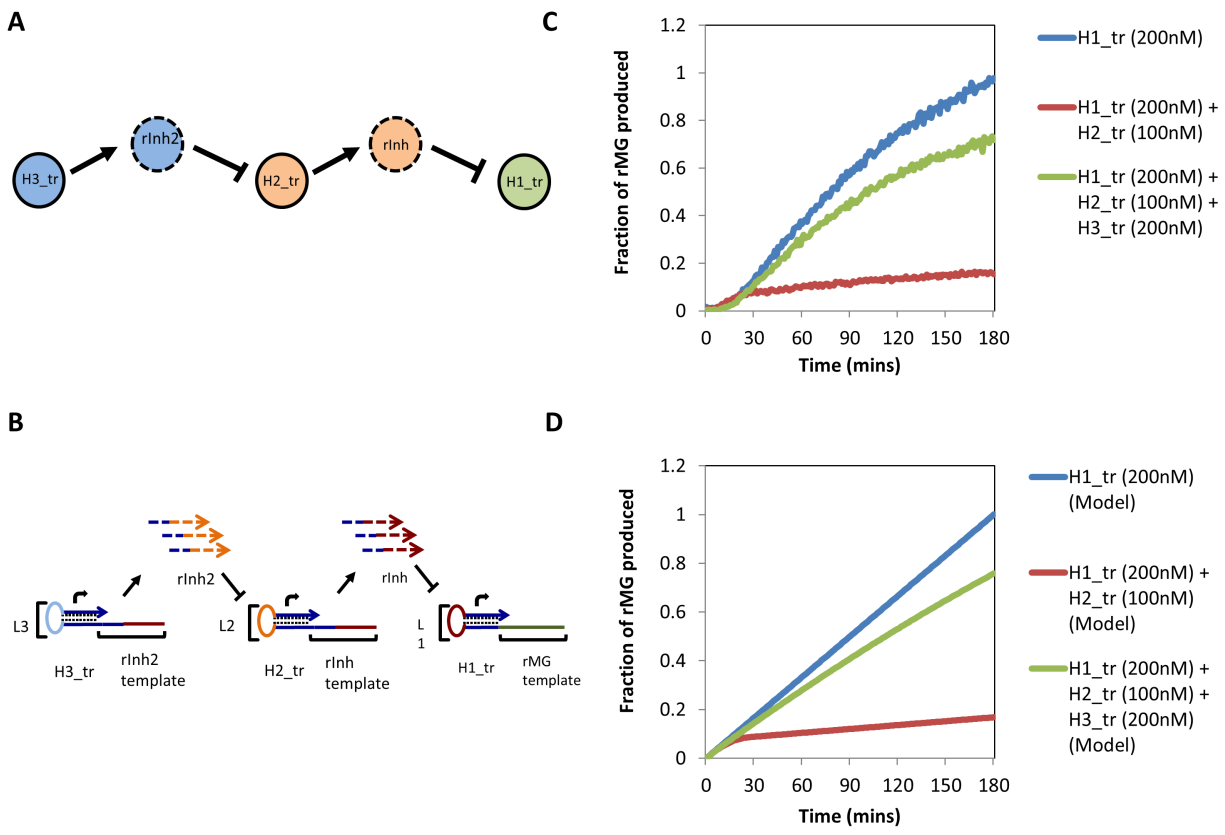

Figure 2.6: Construction and analysis of the double inverter circuit.

A) Nodal representation of the double inverter circuit where the output of H3_tr inhibits the H2_tr which in turn inhibits transcription from H1_tr. B) Schematic demonstrating the double inverter circuit as implemented in the hairpin promoter switch framework. C) Assay to measure the production of rMG in the presence of both H2_tr and H3_tr. D) Simulation dynamics of the modeled reactions involved in the circuit.

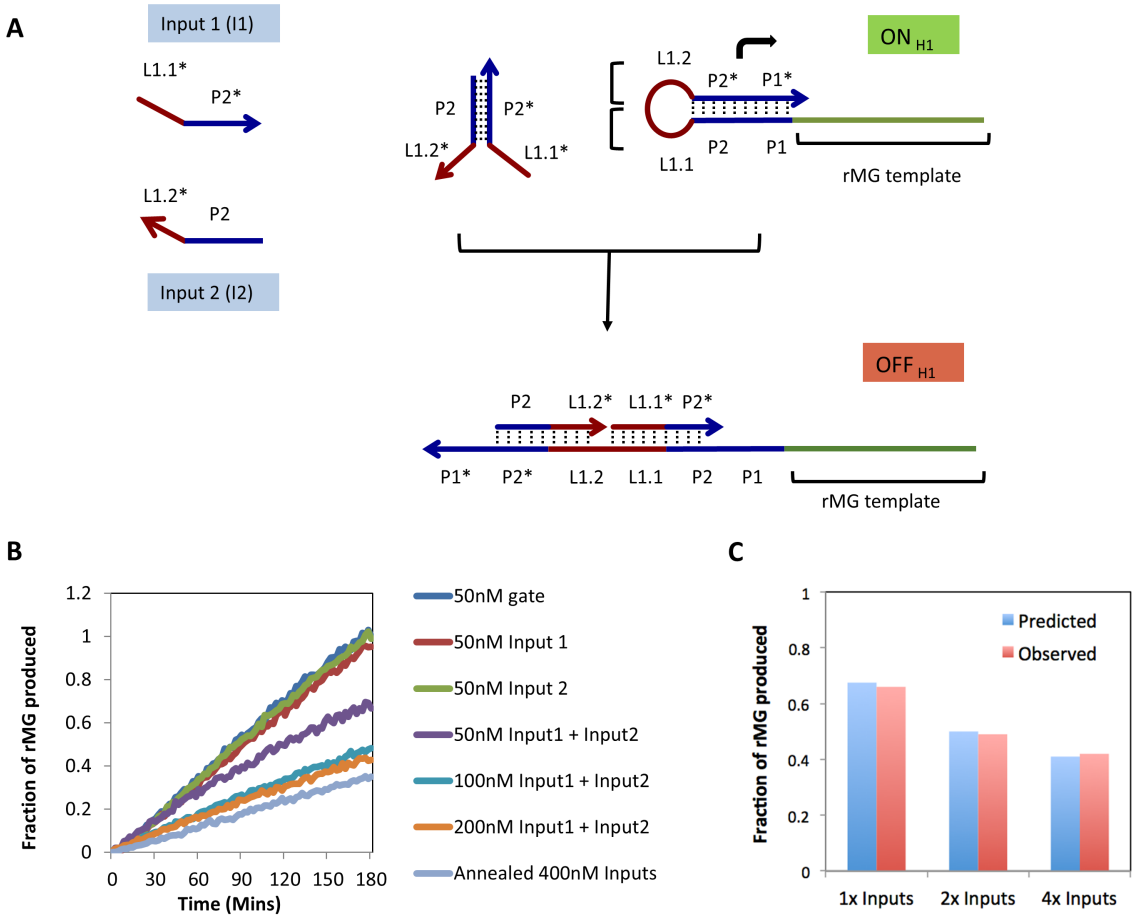

Figure 2.7: Design of NAND gate

A) Schematic showing the simple split inhibitor design where each input has a toehold length of 9nt – L1.1 and L1.2, and a portion of the T7 RNAP promoter (P1 and P2). Cooperative binding of the both inputs leads to the formation of OFF_{H1}. B) Assay of rMG expression after the addition of varying amounts of inputs compared to the annealed OFF_{H1}. C) Comparison of rMG expression from the empirical model developed taking into account the thermodynamic stability of the OFF_{H1} in the presence of the varying amounts of inputs.

For instantaneous inhibition i.e assuming the kinetics of OFF switch formation in the presence of input strands is fast, the concentration of rMG at any time T is

$[rMG]_T = T*ON\_rate*[ON]_{t=T} + T*Leak*[OFF]_{t=T}$

Where $[ON]_{t=T}$ and $[OFF]_{t=T}$ are the respective amounts of the uninhibited and inhibited hairpin template present respectively. Since we assume that the inhibition is instantaneous, relative amounts of [ON]t=T and [OFF]t=T compared to [ON]t=0 can be calculated from equilibrium fractions determined by NUPACK.
ON_rate is the rate of transcription from the ON switch and Leak is the rate of transcription from the OFF switch.

We assume that the leak reaction is a fraction of the ON rate.
Leak = fr*ON_rate where fr = (0,1)

Let F(rMG) = fraction of rMG produced at time T relative to the uninhibited case, i.e $F(rMG) = [rMG]_T / [rMG]_{T, No\ inputs}$

$[rMG]_{T, No\ inputs} = T*ON\_rate*[ON]_{t=T}$

When there are inputs are zero then $[ON]_{t=T} = [ON]_{t=0}$
thus

$$F(rMG) = (T*ON\_rate*[ON]_{t=T} + T*Leak*[OFF]t=T) / T*ON\_rate*[ON]_{t=0}$$

$$F(rMG) = [ON]_{t=T}/[ON]_{t=0} + fr*[OFF]_{t=T}/[ON]_{t=0}$$

where $[ON]_{t=T}/[ON]_{t=0}$ indicates the relative amount of ON template present relative to the initial amount ON switch added.

For the case when the ON is annealed with excess inputs, $[OFF] \sim [ON]_{t=0}$ thus
$F(rMG)_{annealed\ inputs} = fr$
$F(rMG)_{annealed\ inputs}$ is calculated from the data when the NAND gate is annealed with excess inputs.

For the other cases, F(rMG) can be calculated by using NUPACK to calculate $[ON]_{t=T}/[ON]_{t=0}$ and $[OFF]_{t=T}/[ON]_{t=0}$

fr	0.38
$[ON]_T/[ON]_{t=0}$ For 1x Inputs	0.46
$[OFF]_T/[ON]_{t=0}$ For 1x Inputs	0.54
$[ON]_T/[ON]_{t=0}$ For 2x Inputs	0.2
$[OFF]_T/[ON]_{t=0}$ For 2x Inputs	0.8
$[ON]_T/[ON]_{t=0}$ For 4x Inputs	0.04
$[OFF]_T/[ON]_{t=0}$ For 4x Inputs	0.96

Figure 2.8: Empirical model to determine the repression efficiency of NAND gate in the presence of varying amounts of inputs.

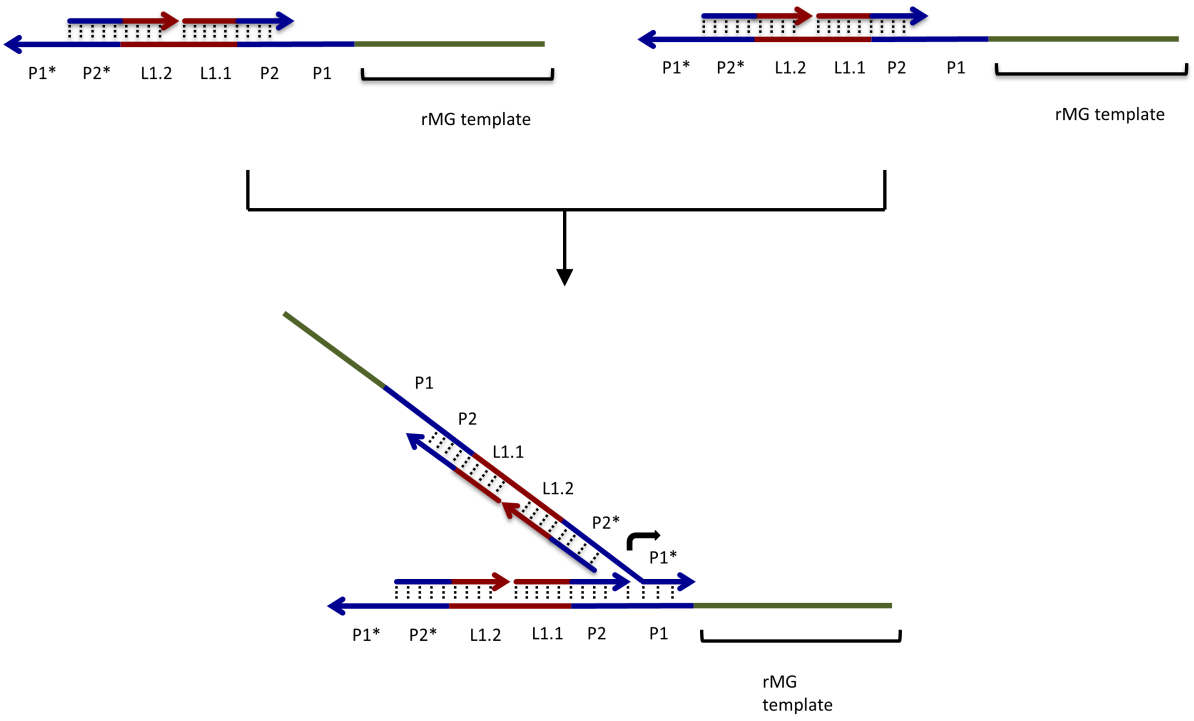

Figure 2.9: Leaky expression of rMG from OFF NAND switch.

Two OFF switches can form an active promoter leading to leaky rMG expression mediated by P1 domain.

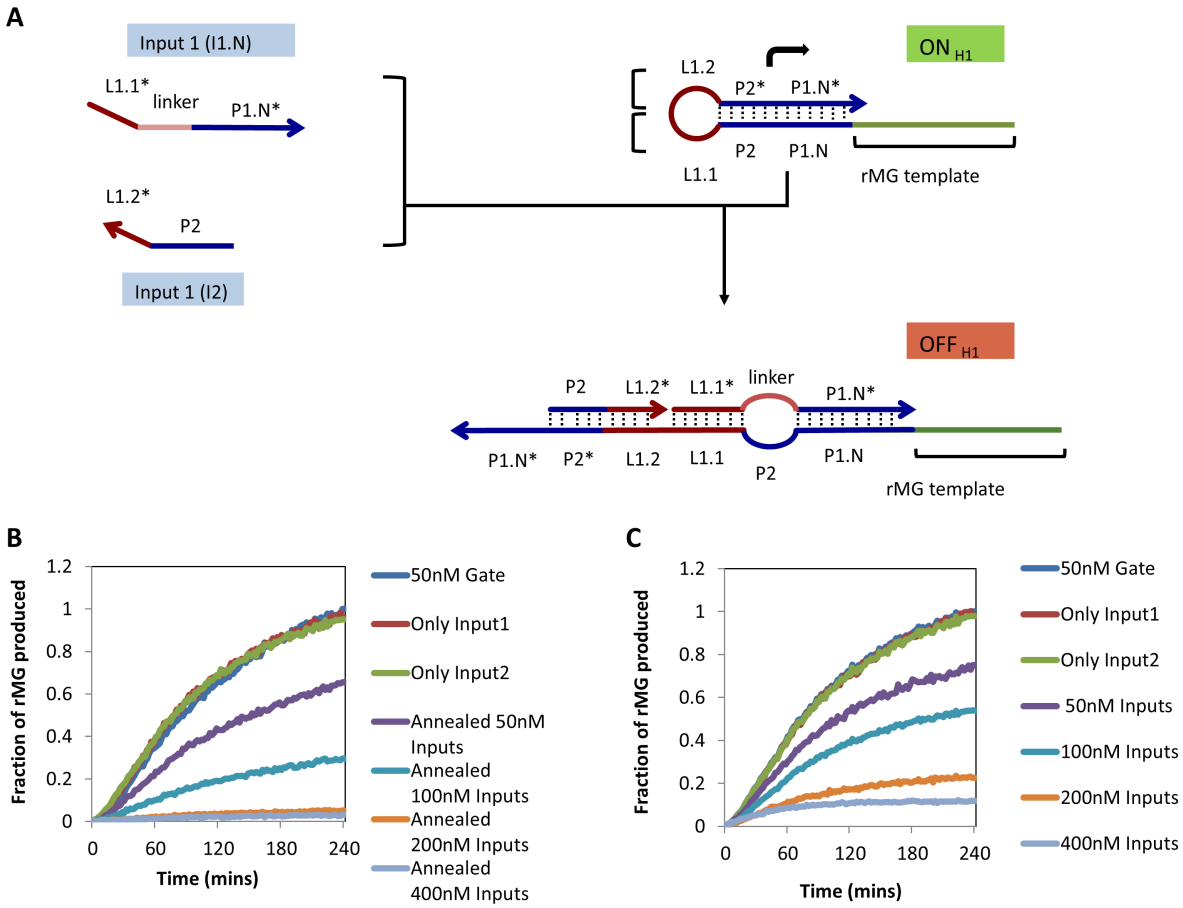

Figure 2.10: Design of the NAND gate with remote toehold design.

a) Remote toehold design for NAND gates. Input I1.N has a remote toehold L1.1* and targets a different region of the T7 RNAP promoter P1.N than in the previous case (P1). b) Assay of rMG expression from annealed OFF_{H1} in the presence of varying amounts of inputs I1 and I2. c) Assay of rMG expression from annealed OFF_{H1} after the addition of varying amounts of inputs I1 and I2 (no annealing).

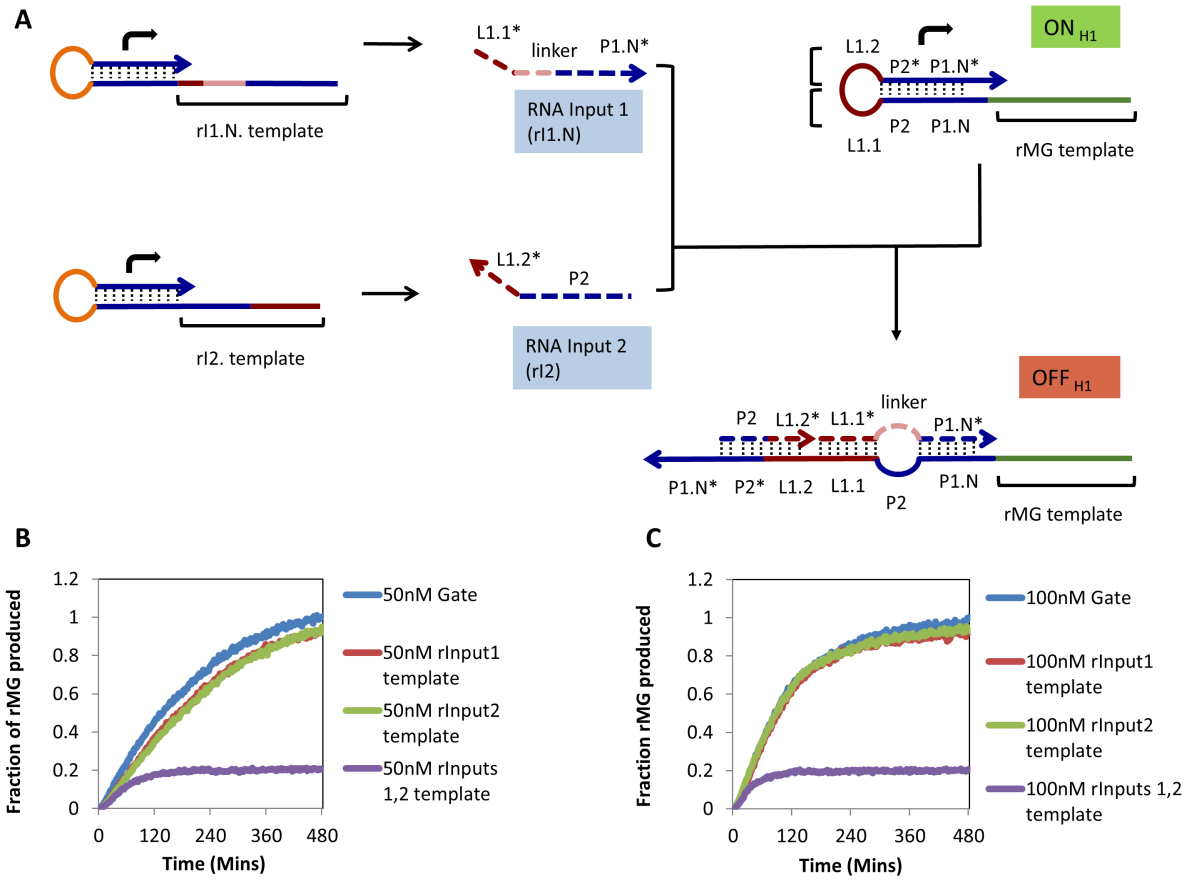

Figure 2.11: Performance of the NAND gate with RNA inputs.

A) Schematic showing the NAND gate design in the presence of RNA inputs (rI1 and rI2) expressed from other hairpin switches. The RNA inputs have the same remote toehold design as for the DNA inputs (Fig 5). b) Performance of 50nM NAND gate in the presence of equivalent amounts of input templates. c) Performance of 100nM NAND gate in the presence of equivalent amounts of input templates.

Table 2.1: List of oligonucleotide sequences used

Name	Sequence
H1_tr	GATCCATTCGTTACCTGGCTCTCGCCAGTCGGGATCCCTATAGTGAGTCGTATTACGGTG TGTAGTAGTAGTTCACGTAATACGACTCACTATAGGG ATC
H2_tr	GGTGTGTAGTAGTAGTTCACGTAATACGACTGATCCCTATAGTGAGTCGTATTACCTCTTACTCCTGATCTGATGGTAATACGACTCACTATAGGGATC
H3_tr	ACCTCTTACTCCTGATCTGATGGTAATACGACTGATCCCTATAGTGAGTCGTATTACGCCATTGACCTAGAAGACTTCGTAATACGACTCACTATAGGGATC
H2_tr_anti	CGTCGTATTACGTGAACTACTACTACACACCGATCCCTATAGTGAGTCGTATTACGCCATTGACCTAGAAGACTTCGTAATACGACTCACTATAGGGATCGGTGTGTAGTAGTAGTTCACGTAATACGAG
I1	ACTACACACCGTAATACGACTC
I2	GAGTCGTATTACGTGAACTACT
I1.N	ACTACACACCGATAAAAGACTCACTATAGGGATC

Chapter 3: Construction of synthetic T7 RNA polymerase expression systems

This chapter is reproduced (with minor modifications) from its initial publication:

Kar, Shaunak and Ellington AD (2018) Construction of synthetic T7 RNA polymerase expression systems *Methods.*

Author contributions:

S.K and A.D.E conceived of the idea. S.K designed and performed the experiments and analyzed the results with input from A.D.E. Both authors wrote the manuscript.

ABSTRACT

T7 RNA polymerase (T7 RNAP) is one of the preferred workhorses for recombinant gene expression, owing in part to its high transcriptional activity and the fact that it has a small (17 base-pair), easily manipulated promoter. Furthermore, the fact that T7 RNAP is largely orthogonal to most hosts enables its use in a wide variety of contexts. However, the high activity of the enzyme also often leads to an increased fitness burden on the host, limiting the predictability of its interactions with and impact on physiology, and potentially leading to mutations to constructs. Here we use a synthetic biology approach to design and characterize a panel of T7 RNAP expression circuits with different modes of regulation that enable the reliable expression of downstream targets under a variety of conditions. First, we describe the construction of a minimal T7 RNAP expression system

that is inducible by a small molecule anhydrotetracycline (aTc), and then characterize a self-limiting T7 RNAP expression circuit that provides better control over the amount of T7 RNAP produced upon induction. Finally, we characterize a so-called T7 RNAP homeostasis circuit that leads to constitutive, continuous, and sub-toxic levels of T7 RNAP. Coupled with previously characterized mutant T7 RNAP promoters *in vitro*, this modular framework can be used to achieve precise and predictable levels of output (sfGFP) *in vivo*. This new framework should now allow modeling and construction of T7 RNAP expression constructs that expand the utility of this enzyme for driving a variety of synthetic circuits and constructs.

INTRODUCTION

Synthetic biology is an engineering discipline in which the modular parts, circuits, and systems can be used to generate devices that have predictable behaviors [36, 113]. Engineered synthetic circuits in prokaryotes often focus on gene regulation through well-characterized inducible transcription factors and computationally designed ribosome binding sites (RBSs) [25, 36, 113]. More recently, there has also been a renaissance in the design and construction of RNA-based tools such as riboswitches and ribonucleoprotein particles such as CRISPR elements to control gene expression [38, 39, 43, 46].

However, while in theory these various parts are modular and orthogonal, in practice introducing synthetic circuits into cells leads to the sharing of essential resources like RNA polymerases (RNAPs) and ribosomes, which can in turn lead to

unpredictable and undesirable outcomes [36, 50]. While translational decoupling of the circuit from the host has proven difficult, transcriptional interference can be avoided by simply utilizing a completely orthogonal 'engine' for transcription than the endogenous RNA polymerase; this also opens the possibility of creating multiple, different orthogonal circuits in parallel using RNA polymerases with different promoter variants [61, 63, 114]. When combined with RNA-based tools that could themselves be produced orthogonally, this engine could be built into a platform for gene expression and regulation that would be both complex and largely outside of the cell's own operating system, potentially exhibiting minimal cross-talk.

Owing to its robust and highly specific activity towards its own promoter, T7 RNA polymerase (T7 RNAP) has been the workhorse for recombinant protein expression for decades [52, 115, 116]. The polymerase is a single subunit and thus can be produced from a single gene. The activity of the polymerase does not rely on any host dependent cofactors, and enables the production of RNA in abundance upon the recognition of its short, 17 base pair (bp) promoter, which is largely orthogonal to *E. coli* and other host RNAPs [52, 115, 116]. This decoupling of transcription from host RNAPs has allowed its use across a wide variety of species [53, 54, 65, 117]. The utility of T7 RNAP as an orthogonal transcription engine also lies in the fact that it is extremely active with rates of transcription 8-fold higher than the *E. coli* RNAP [118, 119]. Moreover, decades of biochemical studies have provided us with a detailed mechanistic understanding of T7 RNAP transcription, including the energetics and kinetics of promoter binding, initiation, elongation, and termination [58, 59, 120, 121]. This has proved to be crucial to the

engineering of novel T7 RNAP variants that recognize orthogonal promoters, split versions of the polymerase that allow construction of resource allocator modules, and mutant T7 RNAP promoters and novel terminators that can tune T7 RNAP associated gene expression [56, 61, 63, 114]. Overall, the use of T7 RNAP for the expression of synthetic circuits is merely a starting place for a broad and expanding toolkit of modular parts that enable precise and multiplex control of gene regulation. As examples, T7 RNAP-mediated expression has found use for expressing circuit components, metabolic pathways, and more recently toehold-based riboregulators [39, 56, 62, 65]. In each instance it has been shown that the decoupling of transcription from the host RNAP led to improved part performance and higher metabolic yields [39, 56, 65].

Unfortunately, due to associated toxic effects possibly due to its increased rate of transcription, the expression of T7 RNAP needs to be finely tuned [56, 63, 65]. Here we describe the construction and analysis of synthetic circuits built from modular parts to control and tune the expression of T7 RNAP, which can then be used as an engine to achieve reliable and predictable expression of downstream genes, pathways, and circuits. The strength of T7 RNAP-mediated expression was compared to known constitutive *E. coli* promoters in order to benchmark the utility of the constructs.

RESULTS

Construction of an aTc-inducible T7 RNAP expression system:

First we designed and built an anhydrotetracycline (aTc)-inducible T7 RNAP expression system, which exhibited no observable leak and allowed for strong/over-

expression of target genes upon the addition of the inducer. TetR was chosen to control the induction as it has proven to be one of the most tightly controlled regulatory, and has been shown to function across species [9, 65]. The T7 RNAP gene was cloned on a medium-low copy number plasmid - p15A/AmpR (10-15 copies/cell) under the control of the $P_{LTetO-1}$ promoter, which is tightly repressed by TetR, which was in turn constitutively expressed from the P_{bla} promoter [9] (**Fig 3.1**). To assay the function of T7 RNAP, a fluorescent reporter (sfGFP; **Table 3.1**) was placed under the control of the wild-type T7 promoter (**Table 3.1**) in a second medium copy number plasmid (pBR322/KanR).

Traditionally, in order to ensure the efficient translation of recombinant mRNAs a strong ribosome binding site (RBS) – AGGAG is used [52]. However expression of T7 RNAP at higher levels contributes to leaky expression even in the absence of induction [52, 115]. In order to mitigate this, weaker ribosomal binding sites (RBSs) were employed to limit the translation of T7 RNAP, an approach that has previously been shown to be effective in cloning and expressing otherwise toxic genes [122]. We chose three RBSs from a set that was previously created by the Anderson lab and deposited in the registry of standard biological parts [49, 123, 124]. The translation strengths of the weaker RBSs were determined relative to that of a strong RBS (AGGAG) whose expression efficiency had been examined previously [16, 17]. The strengths of the weaker RBSs were then further validated in our circuit by measuring the expression of sfGFP after the induction of T7 RNAP with aTc. (**Fig 1b**). Read-through expression of genes that would have disrupted circuit function was minimized

via the introduction of two bacterial terminators (rrnD T1and rnpB T1) with known termination efficiency (>98%, [125]), upstream of the $P_{LTetO-1}$ and the other downstream of the T7 RNAP gene. The use of two unique terminators mitigates the chances of recombination-mediated deletion that otherwise might lead to loss of circuit functionality [126].

For each construct, sfGFP expression was analyzed as a function of the aTc concentration as described in **Methods**. From the aTc dose-response curves it can be seen that the sfGFP signal saturates even at sub-saturating aTc concentration (10ng/mL), which is expected given the strong activity of even weakly expressed T7 RNAP (**Fig 3.1**). At aTc concentrations around 10ng/mL we begin to observe not only strong expression of sfGFP but also growth defects (**Fig 3.1**), which again is consonant with the expression of highly active phage polymerases [56, 63, 65]. For all the assays, the negative control was a circuit expressing an inactive version of the T7 RNAP that included two premature stop codons, as described in [61]. Overall, this system exhibited no observable leak in the absence of any inducer (at least 100-fold that less than maximum signal observed), but the saturation of sfGFP expression even at low induction concentrations provided only limited control over gene expression.

Construction of an inducible and self-limiting T7 RNAP expression system:

With the simple aTc inducible T7 RNAP expression system described in **Methods**, we were able to obtain strong expression of sfGFP with even small amounts of inducer. While such strong expression is well-suited for producing large amounts

of recombinant proteins [52], it is undesirable for consistent and/or regulated production of circuitry. Thus, we designed and constructed a self-limiting T7 RNAP expression system where TetR expression was placed under the control of the T7 promoter (as opposed to being constitutively expressed, **Fig 3.2**). In this system $P_{LtetO-1}$ drives the expression of the T7 RNAP gene, which in turn then expresses its own repressor – TetR is involved in a negative feedback loop (**Fig 3.2**). In order to examine the impact of translation on this auto-regulated circuit, T7 RNAP expression was modulated via three different RBSs as before. When the aTc dose-response curve for the circuit was examined, a linear increase in sfGFP expression was observed with increasing inducer concentration up to 50ng/mL, instead of a saturating signal (**Fig 3.2**). This auto-regulated or self-limiting system should allow for much broader control over the amount of T7 RNAP produced, thus mitigating the burden on the cell's resources relative to the simple inducible circuit (**Fig 3.2**).

Construction of a T7 RNAP homeostasis circuit:

Finally, we designed a system in which the steady state expression of T7 RNAP was largely independent of any host RNAP input. In this system, T7 RNAP not only expresses its own repressor (negative feedback loop) but also controls its own production via its own promoter (a so-called autogene [65, 127] or positive feedback loop, **Fig 3.3**). Coupling the negative feedback loop (via TetR) with the positive feedback loop (via the autogene) should lead to the constant and self-regulated production of T7 RNAP, which can then in turn be used to consistently express the downstream targets.

In previous work [65], the strengths of these feedback loops were tuned by varying T7 RNAP and TetR expression via computationally designed RBS sequences with different relative strengths. Here we have extended the system by balancing the relative strengths of the feedback loops using the previously mentioned weak RBSs to express sub-toxic levels of T7 RNA polymerase, and along with a series of mutant T7 RNA promoters yielded a variety of predictable and reliable steady state expression levels of an output gene (sfGFP). (**Fig 3.3**). The decoupling of the production of T7 RNAP from its host should enable the use of this system across species in a way that is more extensible than attempting to control translation, where the idiosyncracies of RBS strengths are often not known.

As before, we attempted to place T7 RNAP expression under the control of three weak RBS sequences. However, we were unable to clone this system using the strongest RBS amongst the three - RBS3 suggesting that the strength of positive feedback (autogene) led to toxic levels of T7 RNAP produced. The production of sfGFP expression was analyzed as described in **Section 2.4** for the other circuit variants (RBS1 and RBS2) where the amount of sfGFP expressed reflected the relative amount of T7 RNAP expressed from the homeostasis circuit (RBS2 > RBS1). The value of sfGFP/OD_{600} was reasonably constant, suggesting that the production of sfGFP (and by inference, the polymerase) is at steady-state (**Fig 3.3)**. It is important to note that steady state GFP expression was determined after approximately 10 generations as described in **Methods**, reflecting the relative stability of these constructs without any significant growth defects (**Fig 3.3)**.

The level of expression with the homeostasis circuit was compared to that obtained with known *E. coli* constitutive promoters, a strong constitutive promoter (J23100) and a weaker constitutive promoter (J23105), both from the Biobricks collection [49, 123]. The signal from one of the T7 RNAP homeostasis circuit variants (RBS2) is about 40% that of J23100 and 200% that of the J23105, indicating that this configuration is capable of driving consistent gene expression at host-relevant levels. Next we show that using the homeostasis circuit we can tune target gene (sfGFP) expression further by introducing mutant T7 RNAP promoters that have previously characterized strengths *in vitro* [57]. Three different promoters were picked, with strengths relative to wild-type of 0.53, 0.21 and 0.09 (**Table 3.1**, [57]). After performing the assay as described in **Methods** we indeed see that the performance of these promoters is similar to their relative activities observed *in vitro* (**Fig 3.3**). These results suggest that the homeostasis circuit as a whole can be fit to a quantitative model to predict the steady state level of target expression. The ability to quickly parameterize the system with relatively few parts changes is a hallmark of a well-designed synthetic biology construct.

Finally, given the orthogonal behavior of selected T7 RNAP variants [56, 61, 63, 114] we speculated that it should be possible to swap these variants for the wild-type enzyme and promoter while maintaining circuit performance. As a simple proof-of-principle we built and characterized a homeostasis circuit (RBS1) containing one of the orthogonal T7 RNAPs – CGG-R12-KIRV (hereby termed Ortho T7 RNAP) and its corresponding promoter [61, 63, 114]. Steady state analysis shows a predictable

fluorescence output that corresponds to the reported activity of this T7 RNAP variant (determined to be around 0.6-0.7 of the WT T7 RNAP) (**Fig 3.4**). This not only shows that T7 RNAP-based circuitry can be made explicitly modular, but also validates that our modeling with this circuitry is highly predictive.

Modeling the dynamics of the T7 RNAP expression circuit:

The dynamics of the T7 RNAP-based expression systems can be captured with a simple ODE-based model that includes the transcription and translation of three components: T7 RNAP, TetR and sfGFP (**Fig 3.5**) [65]. The dynamics of T7 RNAP mediated transcription was modeled as a three-step process - binding of T7 RNAP to its promoter, initiation and elongation to express a mature transcript. (**Fig 3.5**) where the respective parameters were determined using literature values (**Table 3.2**). The negative feedback mediated by TetR was modeled via binding of the TetR to its operator as described in [16, 17, 65]. The translation strengths of the strong RBS had been calculated before [16, 17] while that of the mutant RBSs were estimated as relative to that of the strong RBS depending on the sfGFP expressed from the simple inducible systems. The half-lives of all 3 mRNAs were assumed to be 2.5 minutes, which is the half-life of the majority of mRNAs in *E. coli* during balanced growth [65, 128]. Similarly, the rate of protein degradation (kdil), was assumed to be due to dilution resulting from cell growth [129, 130], since the proteins did not include any special degradation sites or tags. Growth rates for each construct were calculated from the measured OD_{600} as described in **Section 2.4** and were used as a surrogate measure for dilution. The steady-

state values for sfGFP expression expressed as number of molecules per cell were modeled using a numerical simulation based on the ode45 routine in MATLAB. The steady state analysis from this model predicted the number of sfGFP molecules per cell to be around ~10^5, a value observed previously for similar expression systems [65]. To further validate the model the steady state levels of sfGFP were predicted for previously characterized mutant T7 RNAP promoters that had strengths relative to the wild-type promoter of 0.53, 0.21, and 0.09 [57]. Observed relative levels of sfGFP (fluorescence) were 0.55, 0.17, and 0.07, which were in excellent accord with the prediction (**Fig 3.5**). Moreover, these results demonstrate that promoter strength characterizations performed *in vitro* can be used for accurate modeling *in vivo*, further extending the possibilities of our promoter-tuned system relative to RBS-tuned systems.

DISCUSSION

We show the construction and analysis of a variety of T7 RNAP expression circuits using standardized parts. Overall, these circuits allow induction of T7 RNAP-mediated expression, as well as constant and homeostatic control of expression. Target expression can be predictably tuned using only well-characterized mutant T7 RNAP promoters. The modular construction of this circuit should allow for the rational introduction of other parts, such as different repressors or protein degradation tags, and the adherence to a quantitative model should allow the performance of these parts to be predicted in advance. Overall, the orthogonality and broad host range of both T7 RNAP and TetR should promote even broader and more consistent performance across

multiple hosts, as previously observed for other homeostasis circuitry [65]. Since promoters are more orthogonal to cellular operating systems than translational initiation signals, this promoter-tuned homeostasis circuit should prove especially useful in new bacterial species where promoters and RBS have not yet been fully characterized.

MATERIALS AND METHODS

Plasmids and strains:

p15A and AmpR ; Copy number 10-15 copies/cell

pBR322 (ori- pMB1 and KanR ; Copy number 40-50 copies/cell

E.coli strain – DH10B

Common protocols:

Materials:

- LB broth (Fisher, Cat# BP-9723)
- Agar (Fisher, Cat# BP-1423)
- Antibiotics: Ampicillin (Sigma Cat#A0166), Kanamycin (Sigma Cat#K1377)
- Glycerol (Fisher, Cat# BP-229)
- Calcium Chloride (Sigma, Cat# C3306)

Instruments:

- New Brunswick Innova 44R shaker (Eppendorf, Cat# M1282-0000)
- Allegra X-30-R benchtop centrifuge (Beckman, Cat#B06320)

Media preparation: Bacterial strains were propagated in LB (Fisher) which was prepared by dissolving 25g of LB powder into 1L Millipore water. The media was sterilized by autoclaving for 20 mins. For LB-Agar plates, 13g LB and 9g Agar was added to 600mL water and autoclaved. Appropriate antibiotics (100ug/mL ampicillin, 50ug/mL kanamycin) whenever necessary were added after the mixture cooled downed to less than 55°C. Approximately 25mL of LB-Agar was poured into each culture plate which was subsequently stored at 4°C for a maximum of one month.

Preparation of chemical competent cells: Competent stocks of *E. coli* strains were prepared by inoculating 10mL of LB medium with a small portion of cells scraped from the surface of a frozen glycerol stock not previously exposed to $CaCl_2$ (no antibiotics, use sterile technique). Following incubation overnight at 37°C with agitation, cultures were diluted in 500 ml LB medium and further incubated until reaching an OD_{600} between 0.4 and 0.6. Aliquots of approximately 35 ml were centrifuged at 3400 *g* for 10 minutes at room temperature (Allegra X-22 Centrifuge, Beckman Coulter) and the combined pellets resuspended in 150ml of ice-cold 100 mM $CaCl_2$/10% glycerol solution. Cells were centrifuged again at 3400 *g* for 10 minutes at room temperature (Allegra X-30 Centrifuge, Beckman Coulter) in aliquots of approximately 30 ml and the combined pellets were resuspended in 20mL ice-cold 100 mM $CaCl_2$/10% glycerol

solution. Following incubation on ice for 25 minutes 200 µl aliquots of the cells were snap frozen in liquid nitrogen and stored at -80 °C. For each individual transformation 50uL aliquot was used.

Preparation of electrocompetent cells: Electrocompetent cells were prepared by inoculating 5mL of LB media with cells scraped from the surface of a frozen *E.coli* DH10B glycerol stock (no antibiotics, use sterile technique). Cultures were grown overnight at 37°C in a shaking incubator (250 rpm). The next day, cultures were diluted 1:50 in 25ml and grown until reaching an OD_{600} of ~ 0.8.The entire culture was centrifuged at 3200 *g* for 10 minutes at 4°C, the supernatant was subsequently removed and the remaining cells were resuspended in 25mL ice cold 10% glycerol. This process was repeated two more times after resuspending in 12.5mL and 6mL 10% ice-cold glycerol respectively. After the final wash, the remaining cells were resuspended in 200uL of 10% ice-cold glycerol and aliquots of 50uL were made. Aliquots were stored at -80°C for further use.

Cloning:

Materials:

- Phusion DNA polymerase (NEB, Cat# M0530S)
- Zymoclean Gel DNA recovery kit (Zymo Research, Cat# D4002)
- Gibson Mix (as described in [131])

- QIAprep Spin Miniprep Kit (Qiagen, Cat#27106)
- Oligonucleotides (IDT)

Instruments

- Thermocycler
- Heratherm Oven (Thermo Fisher, Cat# 51028112)
- New Brunswick Innova 44R shaker (Eppendorf, Cat# M1282-0000)
- Table top centrifuge (Eppendorf, Cat# 5418000017)

Protocol:

1. The T7 RNAP gene was obtained from constructs in the lab (Table 1, [61]). This gene was cloned downstream of $P_{LTetO-1}$ with RBSs 1 -3 (Table 1) where the RBS sequences were encoded within the primers itself.

2. The sfGFP was cloned into the pBR322 backbone replacing the T5-lac promoter [114] with the T7 RNAP promoter, the consensus of what are termed as class III promoters ([57], Table 1), The same was done for the constitutive promoters (obtained from biobricks) – J23100 and J23105 thus keeping the 5' UTR region of the mRNA the same.

3. All PCRs were performed using Phusion polymerase (NEB) according to the manufacturer's protocol (25 cycles) where the respective primer annealing temperature was calculated using the online IDT OligoAnalyzer tool.

4. The PCR products were analyzed on a 1% agarose gel, and correct size bands were extracted and purified using the Zymoclean Gel DNA recovery kit.

5. The purified products were assembled using isothermal Gibson assembly with a homology region of 25 nucleotides as described in [131] . The concentration of the template was 50ng and the other products were used in equimolar ratio. For fragments sizes < 1 kb, 5-fold excess product was used in the reaction. The Gibson reaction (20uL) was performed at 50°C for 1 hour and the product was stored -20°C.

6. The Gibson reaction product (1uL) were directly transformed into chemical competent DH10B *E.coli* cells using standard heat shock method, and plated on LB Agar plates with the appropriate antibiotic and incubated overnight at 37°C.

7. Single colonies were picked and grown overnight in 3mL of LB medium with the appropriate antibiotics. The following day the cultures were centrifuged at 3200g for 10 minutes and subsequently minipreped using Qiagen Kit according the manufacturer's protocol.

8. All constructs were sequence verified by Sanger sequencing (UT Austin core facilities).

Assaying:

Materials:

- SOC (New England Biolabs, Cat# B9020S)
- Breatheasy membrane (Sigma, Cat#Z380059)
- 96 well microplates (Corning, Cat#07-200-565)
- Anhydrotetracycline (aTc) (Sigma, Cat# A1200000)

Instruments:

- E.coli pulser (Biorad, Cat# 165-2101)
- M200 plate reader (Tecan)
- Allegra X-30-R benchtop centrifuge (Beckman, Cat#B06320)

Protocol:

1. For all assays, the two plasmids were cotransformed into electrocompetent *E.coli* DH10B cells. Double transformation was done using 0.5uL of each plasmid (~50ng) into 50uL cells using a 1.8kV electroporation protocol on an E.coli

pulser. The transformation was recovered in 500uL SOC and outgrown for 1 hour at 37°C.

2. Only 5uL of the outgrowth were plated on LB Agar plates with appropriate antibiotics and incubated overnight.

3. The next day three colonies were picked (biological replicates) and grown in 2mL of LB with appropriate antibiotics overnight for 18 hours. Cultures were diluted next day 1:100 and grown until they reached early log phase (~ 3 hours, OD ~0.3-0.4).

4. For homeostasis circuits, the cultures were further diluted another 1:200 and 100uL sample was added in each well of the 96 well plate. Plates were covered in Breatheasy membrane to allow for aerobic growth and to prevent evaporation. Growth of the samples was maintained at 37°C and orbital shaking at 244.5 rpm using an amplitude setting of 2.5mm (Tecan M200). The plate was assayed every 10 minutes for both sfGFP fluorescence (excitation 485nm and emission 525nm) and OD_{600} (absorbance at 600nm). Background fluorescence was subtracted from the observed reading (determined by the negative control), and the media specific absorbance was determined using a media only control.

5. The steady state value was calculated during the log phase of cells growth (where the effective dilution of all the proteins occurs).

6. For induction based assays, working concentrations – 1ng/mL, 2ng/mL, 5ng/mL, 10ng/mL, 20ng/mL, 50ng/mL and 100ng/mL were made from aTc stock solution (2ug/uL) in 100% ethanol. For each induction condition, 100uL early log phase cultures were added to 2mL LB with antibiotics and aTc, from which 100uL of sample was assayed.

7. Fluorescence measurements and OD600 absorbance was done the exact same as stated in step 4. For end point measurements, the measured values after 4 hours of assay were used

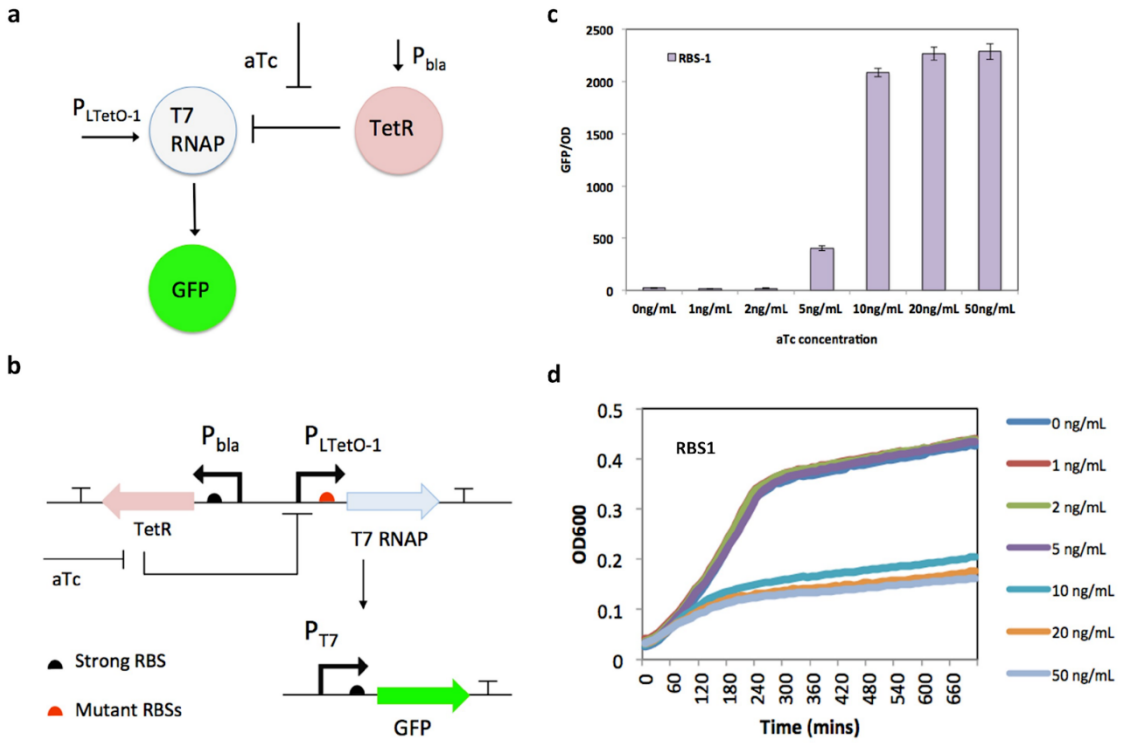

Figure 3.1. An anhydrotetracycline inducible T7 RNAP expression system.

(a) Logic of the circuit where the expression of T7 RNAP is under the control of $P_{Lteto-1}$ promoter where the TetR is constitutively expressed under the P_{bla} promoter (b) Genetic diagram of the construct implemented as two plasmid system. (c) sfGFP production as a function of aTc concentration for T7 RNAP expression under the control of RBS1. (d) Growth curves for cells expressing T7 RNAP (RBS1) and sfGFP after the addition of aTc

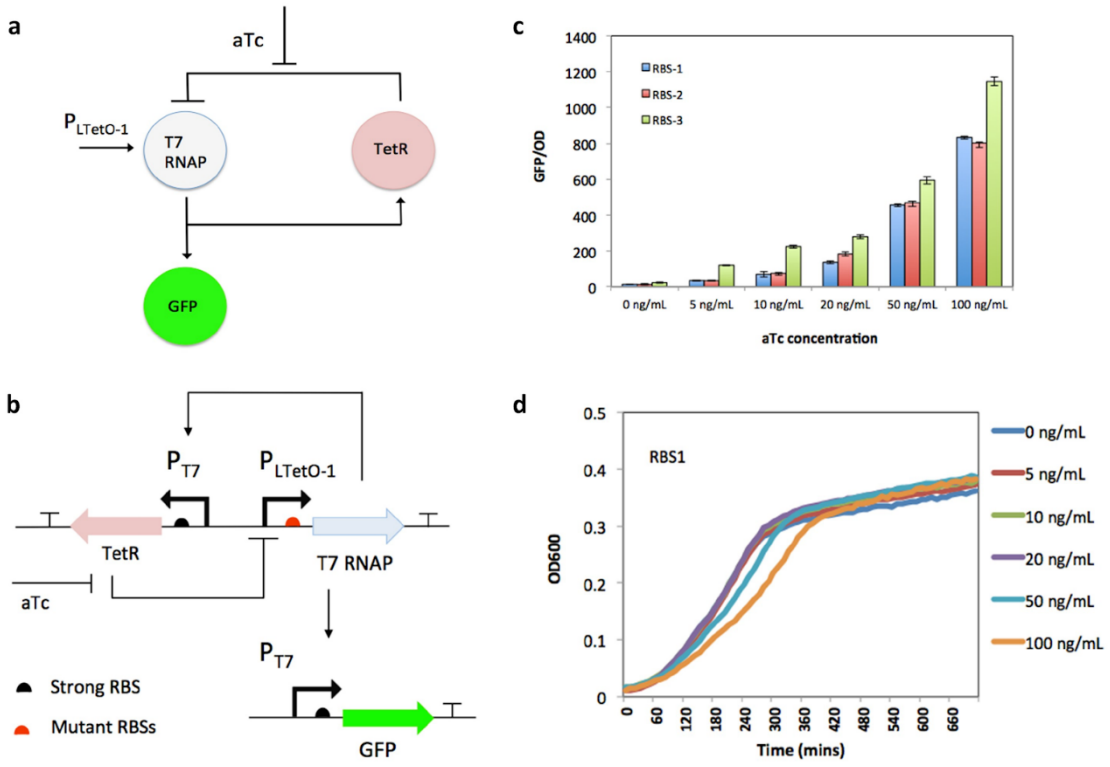

Figure 3.2. Design and characterization of a self-limiting T7 RNAP expression system.

(a) Logic of the circuit where the expression of T7 RNAP is under the control of $P_{Lteto-1}$ promoter and RBSs (1-3) where the TetR expression is under the control of T7 RNAP itself. (b) Genetic diagram of the two plasmid system encoding the self-limiting circuit. (c) sfGFP production from the circuit as a function of aTc concentration with several different RBS strengths. Overall the response is more graded compared to the simple inducible system, in part due to reducing impact of induction on cell growth. (d) Growth curves of cells containing one of circuit variants – RBS1

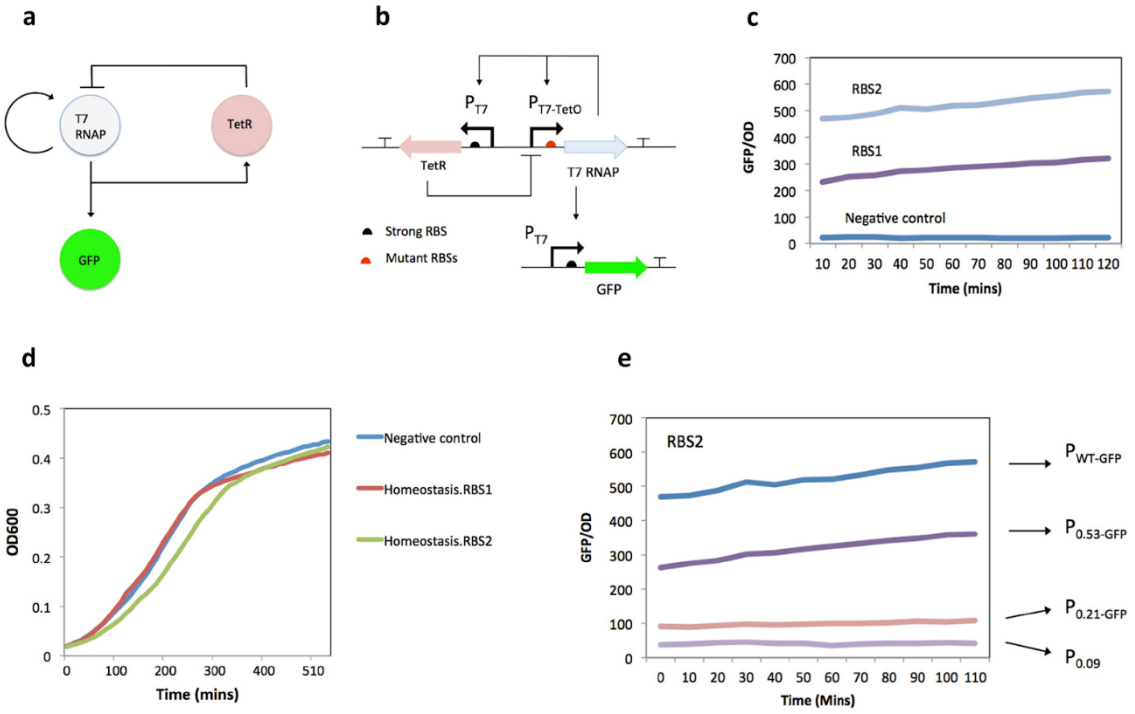

Figure 3.3. Homeostatic control over T7 RNAP production.

(a) Logic of the circuit showing the positive feedback (T7 RNAP autogene) and the coupled negative feedback mediated by the TetR. (b) Genetic diagram of the homeostasis circuit implemented as a two plasmid system. (c) Steady sfGFP production for two homeostasis circuit variants (RBS1 and RBS2). (d) Impact on growth for cells containing functional homeostasis circuit variants (RBS1 and RBS2).

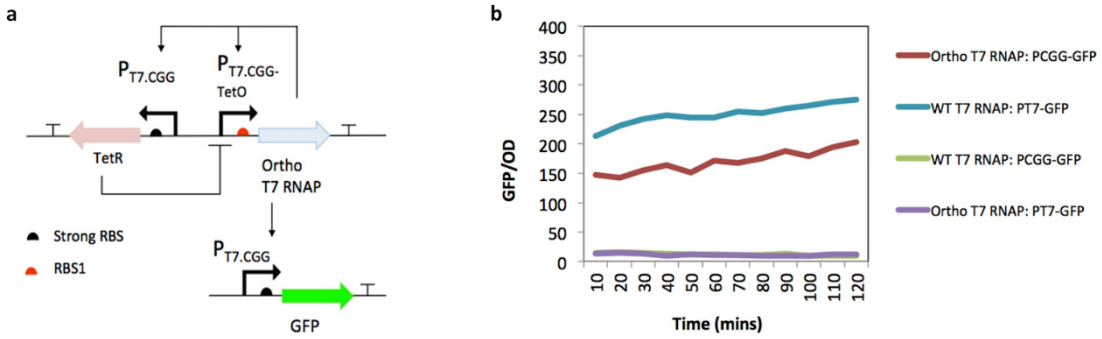

Figure 3.4. Homeostasis circuit utilizing a T7 RNAP variant – CGG-R12-KIRV that recognizes a different T7 RNAP promoter.

(a) Genetic diagram of the homeostasis circuit containing the T7 RNAP variant – CGG-R12-KIRV and the corresponding changes in the promoter sequence. (b) Steady sfGFP production obtained from homeostasis circuit variant utilizing CGG-R12-KIRV under the control of RBS1.

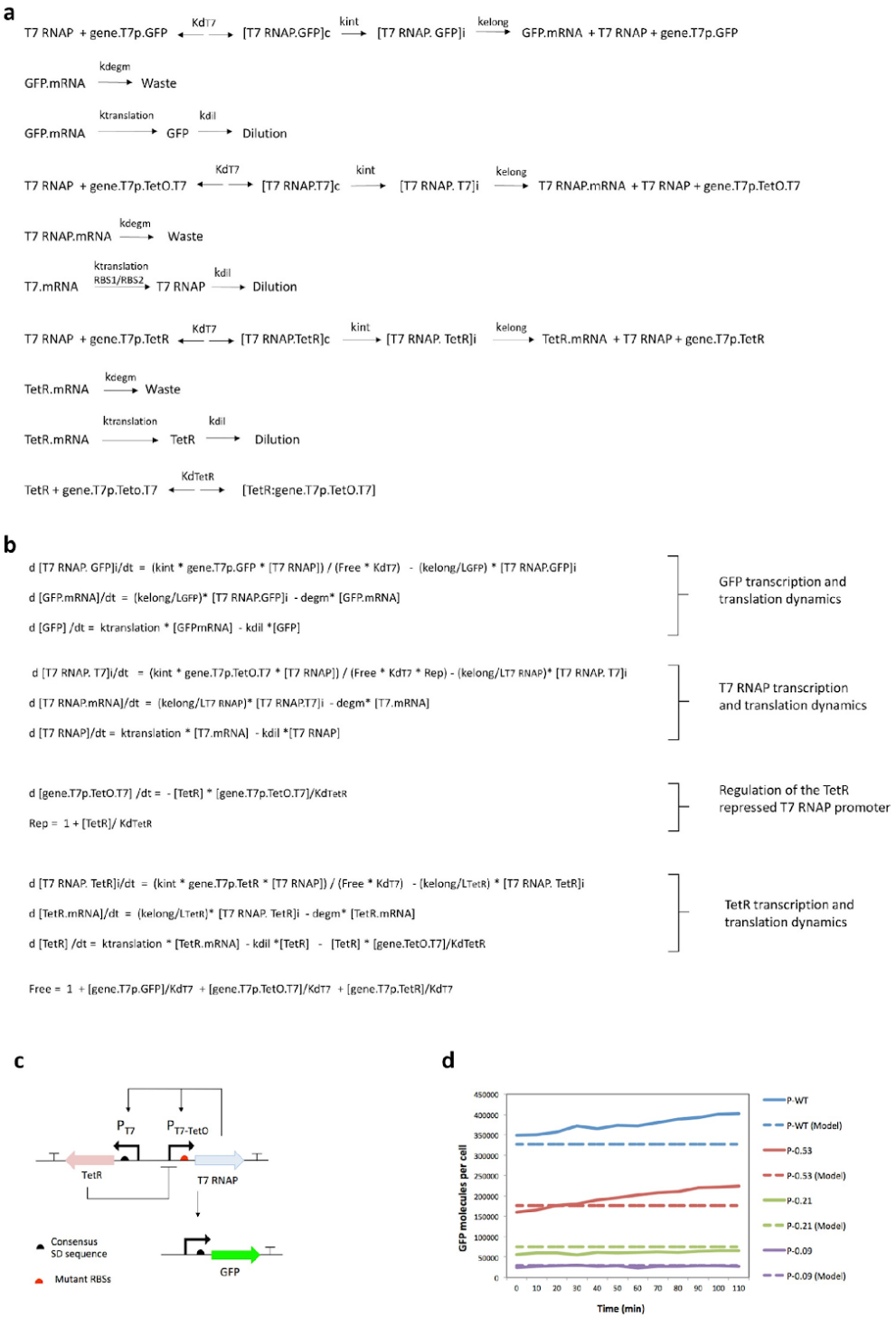
a
T7 RNAP + gene.T7p.GFP ⟷ [T7 RNAP.GFP]c → [T7 RNAP. GFP]i → GFP.mRNA + T7 RNAP + gene.T7p.GFP
GFP.mRNA → Waste
GFP.mRNA → GFP → Dilution
T7 RNAP + gene.T7p.TetO.T7 ⟷ [T7 RNAP.T7]c → [T7 RNAP. T7]i → T7 RNAP.mRNA + T7 RNAP + gene.T7p.TetO.T7
T7 RNAP.mRNA → Waste
T7.mRNA → T7 RNAP → Dilution
T7 RNAP + gene.T7p.TetR ⟷ [T7 RNAP.TetR]c → [T7 RNAP. TetR]i → TetR.mRNA + T7 RNAP + gene.T7p.TetR
TetR.mRNA → Waste
TetR.mRNA → TetR → Dilution
TetR + gene.T7p.Teto.T7 ⟷ [TetR:gene.T7p.TetO.T7]
b
d [T7 RNAP. GFP]i/dt = (kint * gene.T7p.GFP * [T7 RNAP]) / (Free * KdT7) - (kelong/LGFP) * [T7 RNAP.GFP]i
d [GFP.mRNA]/dt = (kelong/LGFP)* [T7 RNAP.GFP]i - degm* [GFP.mRNA]
d [GFP] /dt = ktranslation * [GFPmRNA] - kdil *[GFP]
GFP transcription and translation dynamics
d [T7 RNAP. T7]i/dt = (kint * gene.T7p.TetO.T7 * [T7 RNAP]) / (Free * KdT7 * Rep) - (kelong/LT7 RNAP)* [T7 RNAP. T7]i
d [T7 RNAP.mRNA]/dt = (kelong/LT7 RNAP)* [T7 RNAP.T7]i - degm* [T7.mRNA]
d [T7 RNAP]/dt = ktranslation * [T7.mRNA] - kdil *[T7 RNAP]
T7 RNAP transcription and translation dynamics
d [gene.T7p.TetO.T7] /dt = - [TetR] * [gene.T7p.TetO.T7]/KdTetR
Rep = 1 + [TetR]/ KdTetR
Regulation of the TetR repressed T7 RNAP promoter
d [T7 RNAP. TetR]i/dt = (kint * gene.T7p.TetR * [T7 RNAP]) / (Free * KdT7) - (kelong/LTetR) * [T7 RNAP. TetR]i
d [TetR.mRNA]/dt = (kelong/LTetR)* [T7 RNAP. TetR]i - degm* [TetR.mRNA]
d [TetR] /dt = ktranslation * [TetR.mRNA] - kdil *[TetR] - [TetR] * [gene.TetO.T7]/KdTetR
TetR transcription and translation dynamics
Free = 1 + [gene.T7p.GFP]/KdT7 + [gene.T7p.TetO.T7]/KdT7 + [gene.T7p.TetR]/KdT7
c
P T7
P T7-TetO
TetR
T7 RNAP
Consensus SD sequence
Mutant RBSs
GFP
d
GFP molecules per cell
Time (min)
P-WT
P-WT (Model)
P-0.53
P-0.53 (Model)
P-0.21
P-0.21 (Model)
P-0.09
P-0.09 (Model)

Figure 3.5

Figure 3.5 Modeling of the homeostasis circuit variant

a) Reactions involved in T7 RNAP mediated transcription, translation of three components and regulation of T7 RNAP expression. b) Corresponding differential equations describing the dynamics of the reactions. c) Genetic diagram of the homeostasis circuit implemented as a two plasmid system d) Comparison of the steady state level sfGFP expression obtained from one homeostasis circuit (RBS2) where the sfGFP expression under the control of mutant T7 RNAP promoters (P-WT, P-0.53, P-0.21, P-0.09) as predicted by the model to that obtained from the respective constructs.

Chapter 4: Design and construction of a modular framework for orthogonal and predictable gene expression in prokaryotes

ABSTRACT

Establishing orthogonal transcriptional control over synthetic genetic circuitry is critical in providing robustness, predictability and portability of engineered biological systems. In prokaryotes, this has been achieved with the use of phage orthogonal RNA polymerases (RNAPs) mostly notably T7 RNAP. While T7 RNAP based expression of transgenes has been demonstrated to operate under a wide variety of host contexts, due to the severe toxicity associated with T7 RNAP expression, precise control over T7 RNAP expression is critical in achieving feasible genetic constructs. Here, we first designed and built a universal plasmid architecture composed of modular units for the facile construction of genetic circuitry. Specifically, we utilize this framework to characterize T7 RNAP auto-regulatory systems namely homeostasis circuits. By encoding these T7 RNAP systems in a single broad host range plasmid, we demonstrate how such circuits can be utilized as orthogonal expression systems to achieve robust and predictable gene expression across multiple hosts. We show that this network topology can be exploited for the systematic transfer of transcriptional capacity via the homeostasis circuits across multiple species. As examples, we show this circuit encoded in a single broad host range plasmid can be ported across a number of species – *Pseudomonas putida, Salmonella typhimurium*, and *Serratia*

marcescens, and achieves predictable gene expression based solely on mutant T7 RNAP promoters.

INTRODUCTION

The ability to predictably control and alter cellular phenotypes using engineered gene circuits is a long-standing goal of synthetic biology [10, 12, 13, 36, 132]. Over the last two decades, gene circuits have been designed to perform complex computations such as Boolean logic, analog functions, genetic counters and clocks, memory elements and classifiers [16, 17, 20, 25, 26, 35, 75, 77, 78, 123, 133]. In prokaryotes, such computation is typically encoded as synthetic transcriptional regulatory networks composed of engineered promoters, whose activity can be specifically modulated by transcription factors [22, 24-26]. While such a framework of genetic circuit design has been successful for executing complex functions, the dynamics of circuit behavior are intrinsically coupled to the interactions between the host RNA polymerase (RNAP), engineered promoters and transcription factors[36]. This often results in unwanted interactions with the host regulatory elements and has been a primary source of circuit failure [36, 51]. Such variations in expression become more pronounced when these circuits are transferred across diverse host chassis, has thus limited the characterization of gene circuits predominantly in model organisms [36, 51]. Thus, the transfer of genetic circuits to non-model organisms requires the prior design and construction of a suite of genetic parts and circuit elements which can be a tedious and laborious process [55, 134].

To improve the predictability and portability of gene circuits, orthogonal control of transcription of the circuit elements has been proposed [135, 136]. In prokaryotes, this is achieved by placing the target gene(s) under the control of single subunit phage RNAP which act orthogonally to the host RNAP, most commonly T7 RNAP [56, 114]. By decoupling the transcription of target genes from that of the host RNAP, it effectively insulates their expression against any host RNAP associated regulation thus improving the predictability [56, 61-63, 65]. Owing to its robust activity, T7 RNAP based expression systems have been routinely used for the overexpression of recombinant proteins [52-54]. However due to its expression associated toxicity, precise control over the expression of T7 RNAP is required and has often been established with the use tightly regulated host specific transcription systems [56, 115]. Such stringent dependency on host regulation has limited the broader use of T7 RNAP based systems beyond simple over-expression of recombinant genes. To mitigate this dependency, Salis and colleagues demonstrated that an autoregulated T7 RNAP expression circuit, where both the production of T7 RNAP and regulation was placed under T7 RNAP control can be used to express sub-toxic levels of T7 RNAP in different prokaryotes [65]. This pool of T7 RNAP can then be directed to express reporter gene(s)/ pathways in a host independent manner [65]. More recently, we have extended this framework to demonstrate that precise levels of target gene expression can be achieved in *E.coli* with the use of mutant T7 RNAP promoters (**Chapter 3**) [137]. While these studies demonstrated the feasibility of using T7 RNAP homeostasis circuits as an independent gene expression platform, the rapid design and construction of more T7 RNAP driven

circuits across diverse hosts requires a highly programmable architecture for the rapid assembly of any designed genetic programs.

To this end, we have developed a highly modular and generalizable framework for the design and construction of synthetic gene circuits in prokaryotes termed as Universal Plasmid System (UPS). Using UPS, we have formalized the construction of T7 RNAP homeostasis circuits where all the requisite circuit components are encoded as modular transcription units and can be rapidly assembled into any plasmid backbone of choice. Such parallelized construction greatly accelerates the design-build-test cycle for the prototyping of any synthetic gene circuit across diverse host chassis. As a particular example, we show that T7 RNAP homeostasis circuits composed of three transcriptional units, encoded in a single broad host range vector can be successfully utilized as a highly orthogonal expression platform in prokaryotic species – *Escherichia coli*, *Pseudomonas putida*, *Salmonella typhimurium* and *Serratia marcescens*. Furthermore, using the highly modular nature of the UPS framework we demonstrate how mutant T7 RNAP promoters can assembled to generate circuit variants to precisely and predictably modulate the levels output expression across these diverse host chassis.

RESULTS

Design of Universal Plasmid System (UPS) Architecture

As a starting point, we aimed to standardize the construction of synthetic gene circuits. Episomal plasmids encoding gene circuits have a common architecture and are typically comprised of three major sections – origin(s) of replication which is required for

the propagation in the host species, selectable resistance marker(s) and the synthetic genetic circuit components (**Fig 4.1**). Broadly, the origin of replication and selectable marker determine the plasmid backbone while the recombinant genes/circuit components are cloned into suitable locations into the plasmid backbone. While a wide number of plasmid backbones have been characterized in *E.coli*, the engineering of non-model hosts relies on specific backbones/shuttle vectors. Apart from the plasmid backbone, the gene circuit can also be encoded as modular transcriptional units (TUs) which are further composed of fundamental genetic elements namely promotes, ribosome binding sites (RBS), genes and terminators [25]. To accelerate the design-build-test cycle for the construction and prototyping of genetic circuits, a programmable architecture is required where both plasmid backbones and genetic circuits can be combinatorially assembled from simple modular parts. To this end, we have adapted the highly modular cloning framework based on Type IIS restriction enzymes [28, 138, 139] (**Fig 4.1**). The unique property of these enzymes to generate programmable sticky ends allows the sequential assembly of multiple parts in a single reaction [28, 138, 139]. While a number of strategies have been previously reported, we chose to adapt the optimized version originally developed by Dueber and colleagues for the construction of yeast expression vectors instantiated as the yeast toolkit (YTK) [28]. More recently, the highly generalizable YTK architecture was adapted for the construction of mammalian expression vectors termed as MTK [138]. Thus, a similar framework for the construction of prokaryotic gene circuits will broadly standardize the construction of plasmids across both eukaryotes and prokaryotes.

Briefly, in the UPS framework, each component of the plasmid is encoded as modular units /parts flanked by AarI/BsaI restriction sites followed by unique ligation sites (**Fig 4.1**). First, a plasmid backbone comprising of an origin of replication and selectable marker of choice is assembled along with specific landing pads for the insertion of defined transcriptional units [28] (**Fig 4.1, Table 4.1**). In parallel, gene circuits composed of modular elements such as promoters, ribosomal binding sites, coding sequences and terminators are assembled to form transcriptional units (TUs). Each TU is flanked by appropriate 'connector' sequences that dictate the order of their assembly into the plasmid backbone[28] (**Fig 4.1, Table 4.2**). In its current instantiation, the UPS framework enables the parallel assembly of upto eight transcription units, similar to what was previously described [138].

First, we built well characterized *E.coli* origins along with commonly used selectable markers – KanR, AmpR and CamR as separate parts (**Table 4.1**). All origins were combinatorially assembled with each resistance cassette and successfully transformed in *E.coli*. Next, to determine the variation in gene expression that can result from the change in copy number primarily determined by the specific origin of replication, we cloned a constitutive fluorescent reporter cassette (mScarlet-I) with all the plasmid origins along with two different selectable markers – KanR and CamR (**Fig 4.2**), The relative expression from each origin was consistent with previous studies, and the choice of resistance marker did not affect the expression [140]. (**Fig 4.2**).

Characterization of transcriptional and translational elements in UPS

Now, that we demonstrated the assembly of any plasmid backbone of choice, next we sought to incorporate simple transcriptional (constitutive promoters) and translational control elements (RBSs) and characterize their activities within the UPS framework. Moving forward, we chose the broad host origin (RSF1010) and selectable marker (KanR) as our plasmid backbone for the construction and characterization of all genetic elements and circuits [141]. For transcriptional control, we chose a panel of previously characterized *E.coli* constitutive promoters (SK187-198) that were shown to provide a broad range of expression [49]. Using the UPS framework, these promoters were first cloned and maintained as separate parts. To characterize the expression from these promoters, each element along with a strong translation signal driving a reporter gene (sfGFP) followed by an efficient terminator signal and assembled into the broad host plasmid backbone (**Fig 4.3, Table 4.2**). The plasmids were then transformed into *E.coli* DH10B and assayed for reporter expression. Overall, the panel of promoters modulated the expression of sfGFP two orders of magnitude (**Fig 4.3**). Next, we designed a panel of seven RBS sequences to modulate the expression of levels of sfGFP over 3 orders of magnitude (**Fig 4.3**). Each synthetic RBS was assembled with a strong constitutive promoter J23102 driving the expression sfGFP (**Fig 4.3**). Both the rank order of the RBSs as well as their relative strengths was consistent with the predicted values (**Fig 4.3**). Overall, this demonstrates that the UPS framework can be used to incorporate simple genetic elements to modulate levels of target gene expression both on the levels of transcription and translation.

Design and characterization of UPS encoded T7 RNAP homeostasis circuits

Next, we used the UPS framework for the construction of orthogonal transcriptional control systems specifically T7 RNAP homeostasis circuits. As originally described by Salis and colleagues, T7 RNAP homeostasis circuit is an autoregulatory network where both the production and regulation of T7 RNAP expression is placed under the control of T7 RNAP itself [65]. By modulating the strengths of the feedback loops, the authors showed that varying levels of T7 RNAP can be achieved [65]. However, such control was achieved with the use of host specific translational signals thereby limiting their predictability when transferred across species. More recently, we showed that mutant T7 RNAP promoters can be used to precisely control the expression of the output which can be accurately predicted by a quantitative model in *E.coli* [137]. In both cases the portability of such systems required reengineering of the backbone prior to the transfer in other species. Thus, we sought to simplify the design and construction of these circuits under the UPS framework for improved predictability and portability across diverse host chassis.

First, we formalized the design of T7 RNAP homeostasis circuit, where all circuit components were encoded as three separate transcriptional units composed of modular genetic elements (**Fig 4.4A**). The first and the second TU encodes the expression of T7 RNAP which is designed as a positive feedback loop, and the negative feedback via the expression of TetR placed under T7 RNAP control respectively (**Fig 4.4A**). Finally, the output (sfGFP) also placed under the control of T7 RNAP control is encoded as the third transcriptional unit (**Fig 4.4A**). By encoding the entire circuit as modular transcriptional units, it is possible to systematically combine the genetic elements of each transcriptional

unit to achieve defined levels of output. Furthermore, we show how all the transcriptional units can be seamlessly combined in a single broad host range plasmid, thus greatly improving its portability across diverse host chassis.

Each element of the circuit was designed and cloned as separate parts **Fig 4.4A**). Specifically, to minimize read-through transcription, each T7 RNAP promoter was placed downstream of efficient termination signals. Also, sequence divergent termination signals were used to avoid homology-based recombination that can lead to circuit failure. To enable robust expression of circuit components in each TU, we used RBS sequences that have broad activity across diverse species. Specifically, for the expression of T7 RNAP we chose strong translation signals that contain conserved sequence motifs to achieve high levels of output expression (**Fig 4.4A, Table 4.2**). For the expression of the output, we used the RBS (SK136) that showed the highest level of sfGFP expression (**Fig 4.3**). Overall, we designed three versions of the circuit differing only in the RBS controlling T7 RNAP expression (SK249, 250 and 252) and assembled in a single broad host range vector (RSF1010-KanR). All T7 RNAP homeostasis circuits were successfully transformed in *E.coli* DH10B cells and assayed for reporter expression. All circuit variants demonstrated high levels of output expression (200-300 fold) while the leaky expression from the circuits in the absence of T7 RNAP was negligible (**Fig 4.4B**). It is important to note that the output levels from T7 RNAP homeostasis circuit were similar to those obtained from strong constitutive promoters previously described in (**Fig 4.3**).

As previously described the dynamics of this T7 RNAP homeostasis circuit can be modeled using a simple ODE based model describing the expression of T7 RNAP, TetR

and sfGFP [137] (**Fig 4.5**). As suggested by the model, the levels of the output expression levels can be precisely controlled and predicted using the mutant T7 RNAP promoters. We chose a set of six mutant promoters predicted to modulate the output levels over 20-fold. The specific mutations for each promoter variant were cloned as modular parts and combined with the previously transcription units (using SK252 as the wild-type template) to form complete T7 RNAP homeostasis circuit variants – SK267, SK268, SK269, SK270, SK271 and SK272 (**Fig 4.6**). While the absolute levels of output expression depend on a variety of additional factors such as mRNA decay rate, protein degradation rates and translation rates, the relative strengths of expression from the mutant T7 RNAP promoters can be derived by using an internal reference – SK252. This should also allow the comparison of performance of these circuits across different species beyond *E.coli.* Also, it has previously been demonstrated that using an internal *in vivo* standard to quantify activity of promoters leads to improved consistency and accuracy [49]. All circuit variants were transformed in *E.coli* DH10B and the sfGFP levels were characterized. The relative levels of expression from each homeostasis circuit (SK267-SK272 relative to SK252) showed nearly identical levels to that predicted by the model (**Fig 4.6**). This demonstrates how the orthogonal interaction of between T7 RNAP and its promoter can be effectively used to achieve predictable levels of expression in a host independent manner dictated solely by mutant T7 RNAP promoters. Apart from the mutant T7 RNAP promoters, the output levels can be also modulated with the designed translational signals to enable further fine-tuning of output levels (**Fig 4.7**). We incorporated the previously designed RBS panel to control output expression in the T7 RNAP homeostasis circuit (**Fig 4.7**). Thus, overall

T7 RNAP homeostasis circuits can be broadly understood as an orthogonal constitutive gene expression platform where both transcription and translation strengths can be independently controlled.

After we validated our T7 RNAP homeostasis circuit expression in *E.coli* DH10B, we wanted to characterize the expression levels from T7 RNAP homeostasis circuits, in two other *E.coli* strains - MG1655 and Nissle. The latter is of particular relevance since it is a widely used probiotic strain [142]. For both strains, we assayed the expression levels of circuit variants under the control of T7 RNAP promoter variants (SK252, SK267-S272). The relative strengths of expression from these circuit variants showed identical behavior in both strains as predicted by the model while negligible leak was observed in the absence of T7 RNAP (**Fig 4.8**). In addition, for both strains, we characterized the activity of the panel of constitutive promoters previously described in **Fig 4.3**. As expected, the levels of output obtained using constitutive promoters was similar to that generated by T7 RNAP homeostasis circuits (**Fig 4.9**). Thus, this demonstrates that a single broad host range plasmid T7 RNAP homeostasis circuit can be transferred across different *E.coli* strains to obtain predictable levels of expression dependent on T7 RNAP.

Cross-species activity of T7 RNAP homeostasis circuits

Next, we sought to characterize the performance of T7 RNAP homeostasis circuits across diverse prokaryotic host chassis. As mentioned previously, all circuit components were encoded in a single broad host range plasmid which enabled the easy transfer of circuits without any need for additional engineering. We chose three different species of

industrial and therapeutic relevance – *Pseudomonas putida, Salmonella typhimurium*, and *Serratia marcescens* to characterize the performance of T7 RNAP homeostasis circuits. *P.putida* is a well characterized microbe which has been extensively used for industrial production of a wide variety of chemicals as well as being engineered for bioremediation applications [143]. *S.typhimurium* has been demonstrated to selectively colonize the hypoxic conditions, such as those prevalent in the tumor microenvironments and thus serves as an ideal chassis for the selective delivery of therapeutic payloads to these environments [144]. *S.marcescens* is an opportunistic pathogen, and the particular strain N10A28 was first isolated was first isolated from the gut of honey bees and has since been used as part of the bee gut microbiome engineering[141].

First, the panel of constitutive promoters were (SK187-199) were transformed in all species and the expression of sfGFP levels were determined (**Fig 4.10**). As expected, the strengths of these promoters varied across the three species likely due to the differences in the RNAP binding. Next the T7 RNAP homeostasis circuit (SK252) was transformed in the three species and in all cases the circuits demonstrated high levels of output (200-300 fold) (**Fig 4.11**). Remarkably, no significant leak observed in any species, thus suggesting that this one plasmid design can serve as an orthogonal expression platform without any need for additional insulation (**Fig 4.11**). Also, in each species, T7 RNAP mediated expression resulted in similar output levels to that obtained from strong constitutive promoters (**Fig 4.10**). Next, we hypothesized that owing to the orthogonal interaction of T7 RNAP and its promoter, the relative levels of expression obtained of the mutant T7 RNAP promoters driving output expression characterized in *E.coli* will also be consistent

across species. We transformed the panel of circuit variants (SK267-272) in all the species, the relative levels of expression compared to SK252 was characterized. The determined levels were nearly identical as predicted by the model (**Fig 4.11**). Thus, this demonstrates that T7 RNAP homeostasis circuits can be utilized to precisely modulate and accurately predict the output levels across diverse prokaryotic chassis.

DISCUSSION

In this work, we have used a highly modular cloning framework to simplify the design and characterization of orthogonal transcription control systems driven solely by T7 RNAP. The UPS architecture described here can seamlessly be adapted for the construction of arbitrary genetic program beyond T7 RNAP homeostasis circuitry. While in its current instantiation, T7 RNAP homeostasis circuits essentially act as an orthogonal constitutive gene expression platform, the construction of more complex transcriptionally controlled genetic circuits will require additional regulation of T7 RNAP promoters. In prokaryotes, precise control over the activity of synthetic promoters can be achieved by systematically placing transcription factor binding sites (operators) along with host RNAP recognition sites via the sigma factor binding sites to generate the optimal response function. In a similar manner, the operators of allosteric transcription factors, specifically repressors can be used to regulate the activity of a T7 RNAP promoter. By coupling the expression of both the regulator and the target promoter under T7 RNAP control, more complex transcriptionally regulated circuits can be designed and characterized driven solely by T7 RNAP. Furthermore, the response function of such promoters, can Such orthogonal control

systems based regulated T7 RNAP promoters should enable the construction of more complex circuits, whose dynamic behavior will be conserved across species.

MATERIALS AND METHODS

Plasmid design and construction

All oligonucleotides and gblocks were obtained from Integrated DNA Technologies (IDT) unless otherwise stated. For the construction of each genetic element namely promoters, RBSSs, coding sequences and terminators, first they were checked for restriction sites for the following enzymes – BsmBI, BsaI, NotI and AarI. The restriction sites in the coding sequences were removed by the use of synonymous codons while the other elements did not contain any of these restriction sites. The list of parts and constructs are provided in **Tables 4.1** and **4.2**. The part plasmids were cloned into a common vector where each genetic element is flanked by Bsa1 restriction sites followed by appropriate overhangs (**Table 4.1**). For the assembly of both single TU or multi-TU, the following procedure was used: 10 fmol of backbone plasmid and 20 fmol of parts/TUs were used in a 10uL reaction with 1ul of 10x T4 ligase buffer along with 100 units of BsaI-v2 (single TU) or Esp3I (multi-TU or parts) and 100 units of T7 DNA ligase. The cycling protocol used is: 24 cycles of 2 min at 37°C (for digestion) and 3 min at 16°C (for ligation) followed by a final digestion step at 37°C for 30min and the enzymes were heat inactivated 80°C for 20 min. All constructs were transformed into DH10B cells, grown at 37°C using standard chemical transformation procedures. The colonies that lack fluorescence were inoculated

and plasmids were extracted using Qiagen Miniprep kit according to the manufacturer's instructions. Plasmids were maintained as the following antibiotics kanamycin (50ug/mL), chloramphenicol (34ug/mL) and carbenicillin (100ug/mL) wherever required. The plasmids were sequence verified by Sanger sequencing (UT Austin Genomic Sequencing and Analysis Facility). The following enzymes were used for the assemblies – BsaI-v2 (NEB #R3733S), Esp3I (NEB #R0734S) and T7 DNA ligase (NEB #M0318S).

Bacterial strains and culture conditions

E.coli DH10B was used for cloning and initial assay purposes. *E.coli* Nissle was kind gift from the Davies lab (UT Austin). *S.typhimurium* strain ELH1301was a kind gift by the Shapiro lab (Caltech), *S.marcascens* N10A28 was a kind gift by the Barrick lab (UT Austin) and *P. putida* KT2440 was obtained from ATCC (#47054). All bacterial strains were grown in LB at 37°C except Pseudomonas putida which was grown at 30°C. For all strains, the plasmids were maintained using kanamycin (50ug/mL). Electrocompetent cells were prepared by inoculating 2mL of LB media. Cultures were grown overnight at 37°C in a shaking incubator (250 rpm). The next day, cultures were diluted 1:50 in 25ml and grown until reaching an OD_{600} of ~ 0.8. The entire culture was centrifuged at 3200 *g* for 10 minutes at 4°C, the supernatant was subsequently removed, and the remaining cells were resuspended in 25mL ice cold 10% glycerol. This process was repeated two more times after resuspending in 12.5mL and 6mL 10% ice-cold glycerol respectively. After the final wash, the remaining cells were resuspended in 1mL of 10% ice-cold glycerol and aliquots of 50uL were made. Aliquots were stored at -80°C for further use.

Fluorescence assay and measurement

For all assays, the single plasmid (100ng) was transformed into 50uL electrocompetent cells. using 1.8kV using an E.coli pulser (Biorad). The transformation was recovered in 500uL SOC and outgrown for 1 hour at 37°C (and 30°C for *P.putida)* and plated on selective antibiotic plates. The next day, for each construct, three colonies were picked (biological replicates) and grown in 500uL LB with appropriate antibiotics in a 96-well deep well plate (Axygen) overnight for 18 hours with 900 rpm shaking. Cultures were diluted next day 1:100 and grown until they reached early log phase (~ 3 hours), the cultures were further diluted another 1:100 and 100uL sample was added in each well of the 96 well plate assay plate (Corning). Plates were covered in Breatheasy membrane to allow for aerobic growth and to prevent evaporation. Growth of the samples was maintained at 37°C (30°C for *P.putida)* and orbital shaking at 432 rpm using an amplitude setting of 1mm (Tecan M200). The plate was assayed every 20 minutes for both sfGFP fluorescence (excitation 485nm and emission 515nm) and OD_{600} (absorbance at 600nm). Background fluorescence was subtracted from the observed reading (determined by the negative control), and the media specific absorbance was determined using a media only control.

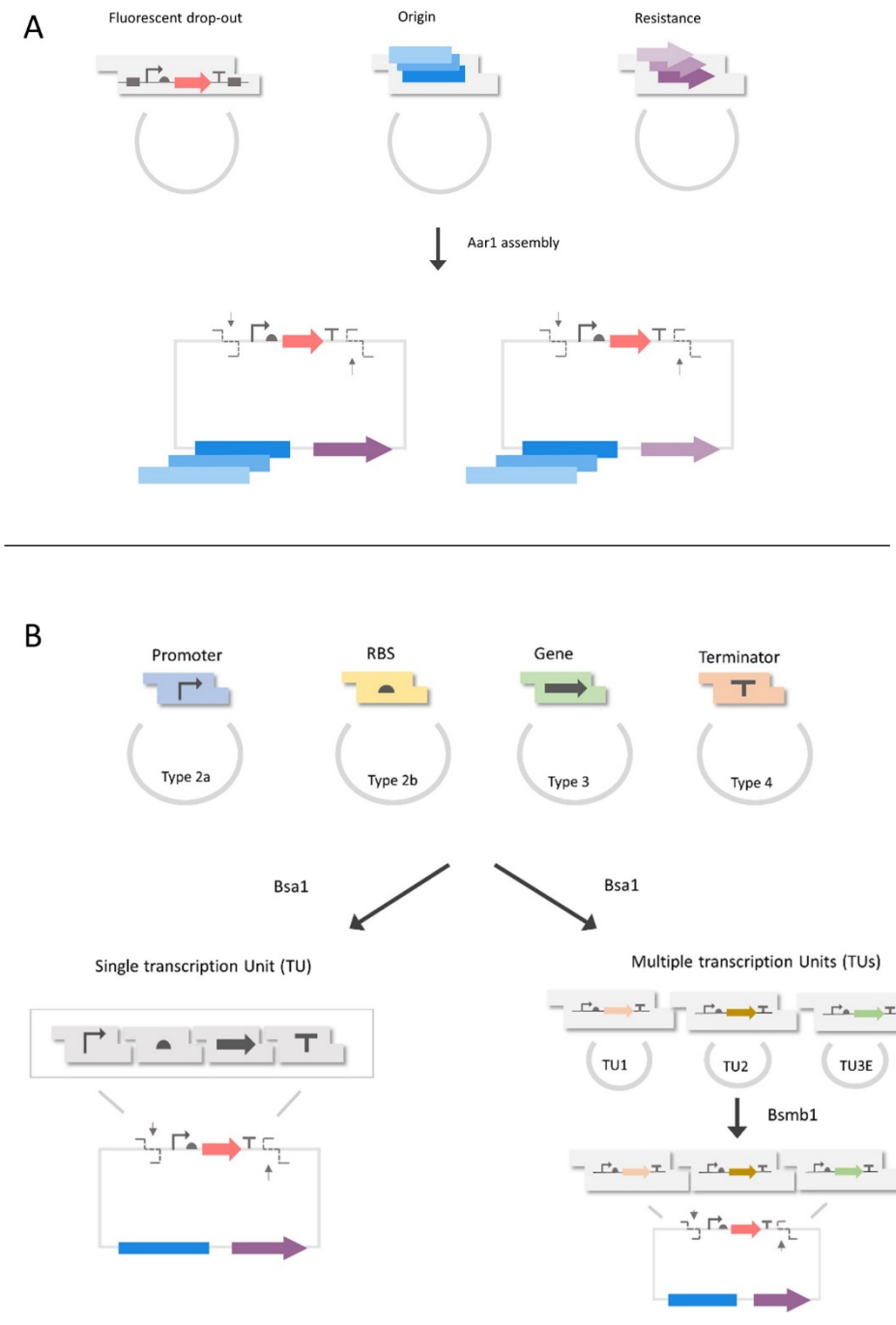
A
Fluorescent drop-out
Origin
Resistance
Aar1 assembly
B
Promoter
RBS
Gene
Terminator
Type 2a
Type 2b
Type 3
Type 4
Bsa1
Bsa1
Single transcription Unit (TU)
Multiple transcription Units (TUs)
TU1
TU2
TU3E
Bsmb1

Figure 4.1

Figure 4.1: Design of genetic constructs with the UPS architecture

A. Assembly of plasmid backbone comprised of origin of replication, resistance marker and a fluorescent drop-out cassette flanked by AarI sites. B. Assembly of genetic circuit encoded as modular parts – promoters , RBSs , coding sequences and terminators. Circuits can be assembled either as a single transcriptional unit (Bsa1) or as multiple transcriptional units (BsmBI) where each transcriptional unit is assembled in parallel and combined togther into the final plasmid backbone of choice.

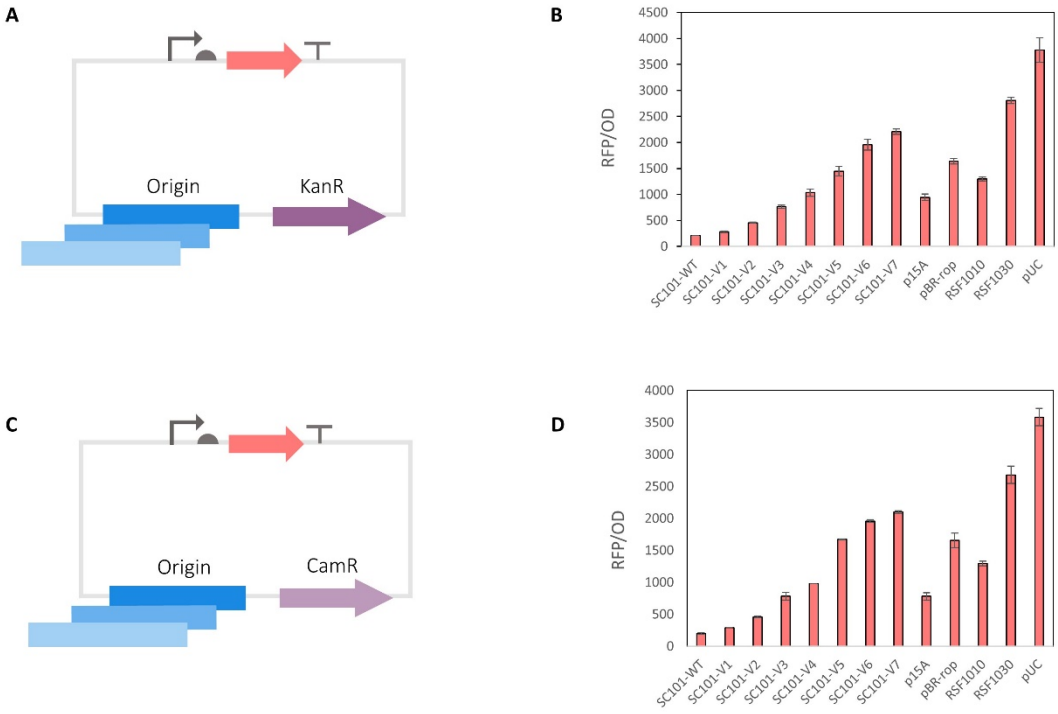

Figure 4.2: Assay of a constitutively driven mScarlet-I to compare the variation in copy number.

Plasmid backbones with different origin of replication and resistance markers were assembled. All the origins shown were assembled with two different resistance markers – KanR and CamR and the fluorescent expression was assayed.

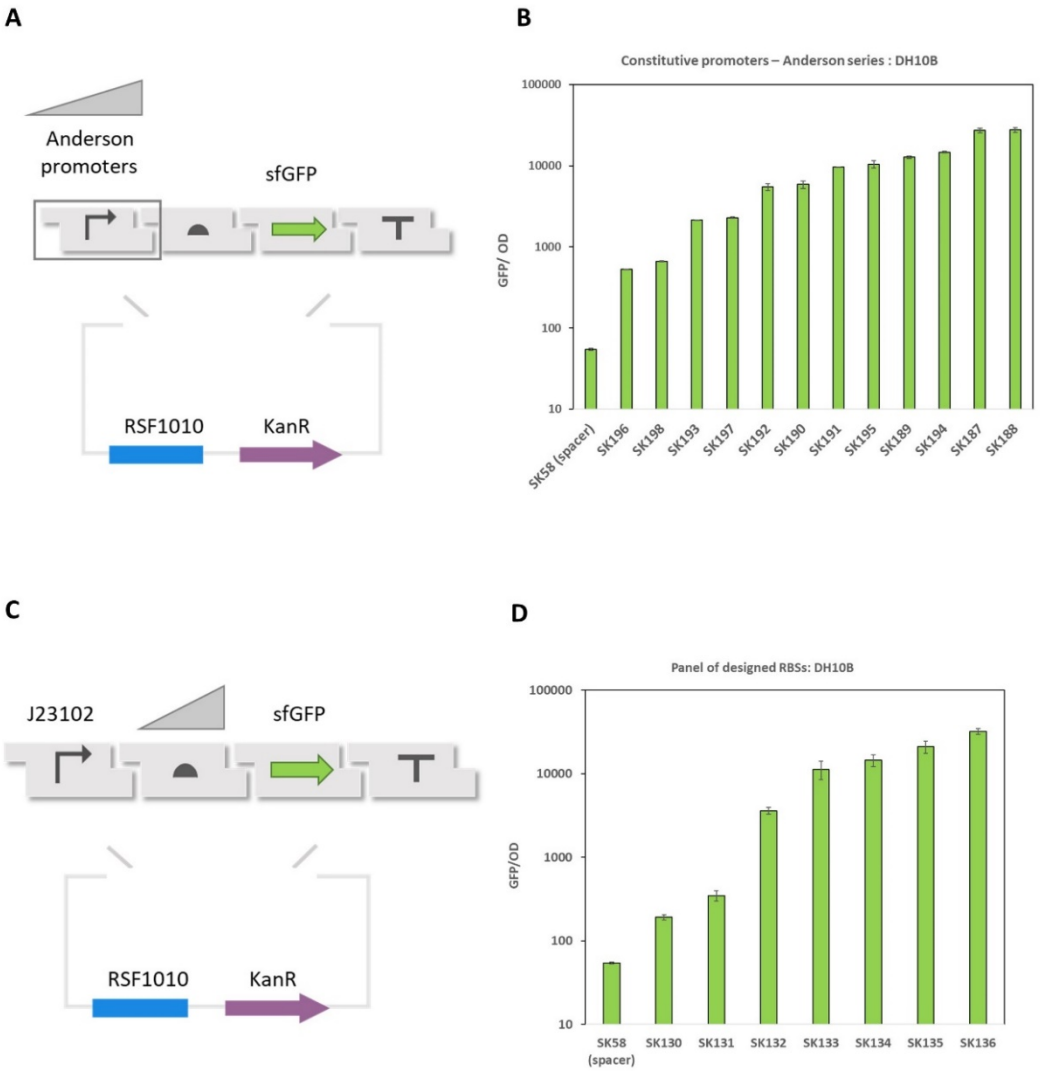

Figure 4.3: Design and characterization of transcriptional control (constitutive promoters) and translational control elements (RBSs)

A. A panel of constitutive promoters was assembled to drive the expression of the output (sfGFP) under the control of a strong translational signal into a broad host range plasmid (RSF1010 and KanR). **B**. Characterization of GFP expression from the panel of promoters. **C**. A panel of seven RBSs were designed to regulate the levels of sfGFP production. **D**. Characterization of sfGFP levels from the panel of RBSs.

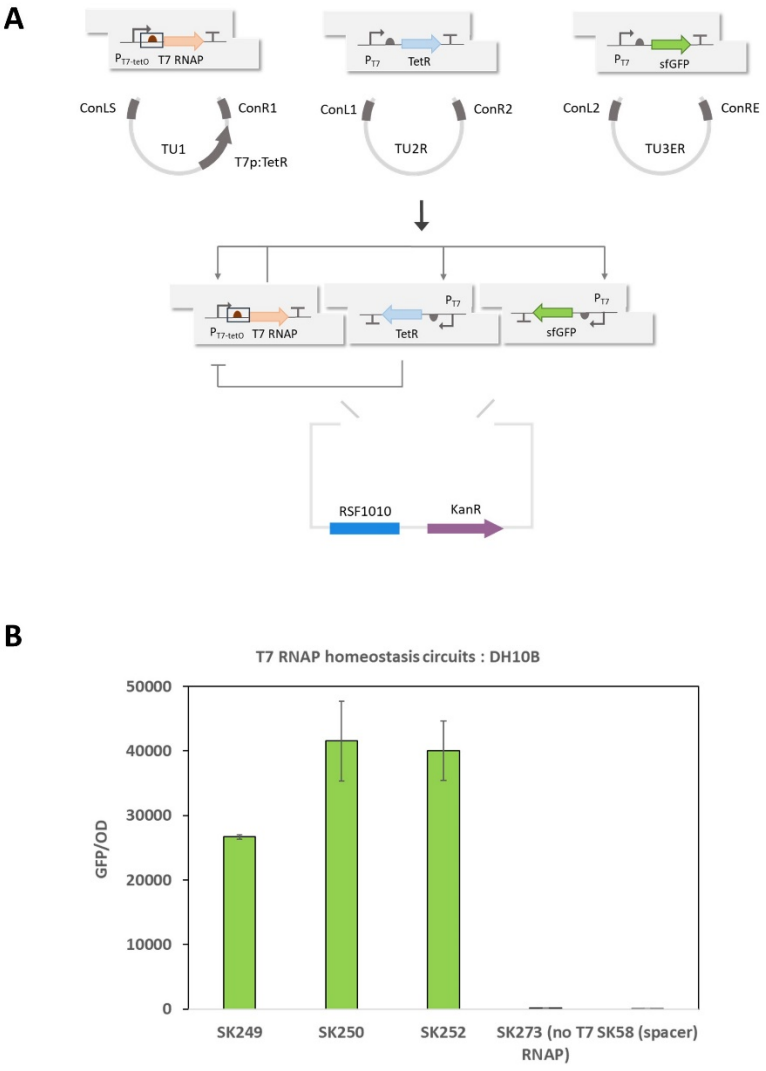

Figure 4.4: Design and characterization of T7 RNAP homeostasis circuit using the UPS architecture

A. Modular design of the T7 RNAP homeostasis circuits encoded as three separate transcriptional units, where each TU is maintained as separate plasmids. Three broad host RBSs were used to drive the expression of T7 RNAP. All the TUs were then assembled into the broad host range plasmid (RSF1010 and KanR) **B.** All three constructs along with SK58 (spacer) and SK273 (no T7 RNAP) were transformed in DH10B and assayed for GFP expression.

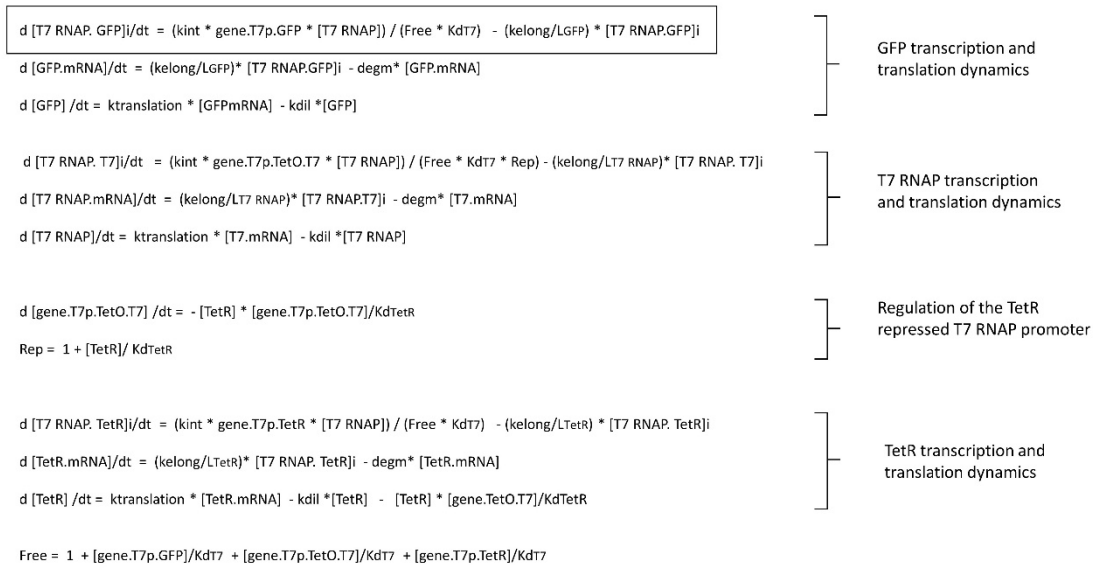

Figure 4.5: Model describing the dynamics of the T7 RNAP homeostasis circuits.

The model describing the dynamics of the T7 RNAP homeostasis circuits has been adapted from that described in Chapter 3. To quantify the levels of the output expressed from mutant T7 RNAP promoters, the highlighted (box) ODE was changed and the levels were normalized to SK252.

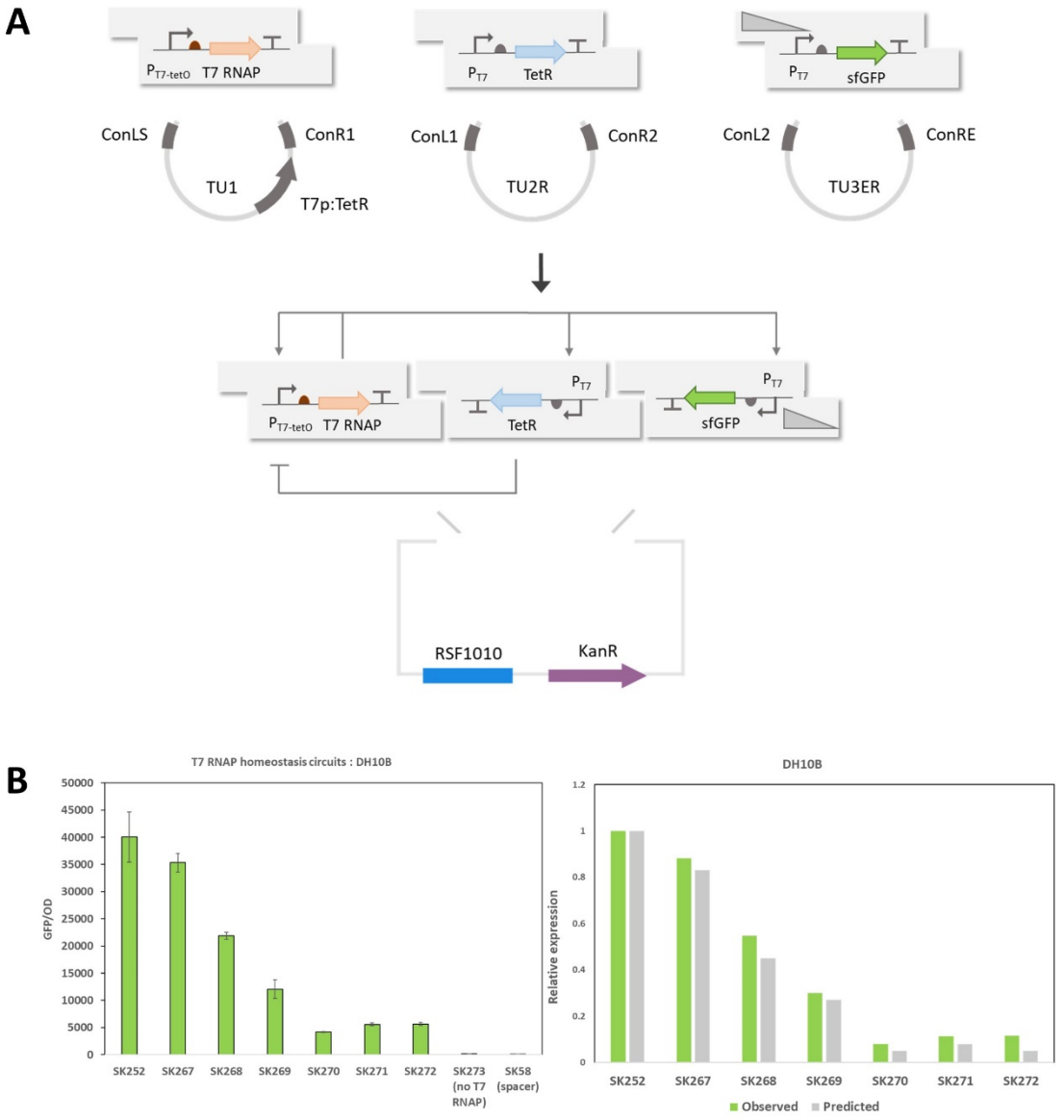

Figure 4.6: Comparison of output expression from T7 RNAP homeostasis circuit to that predicted by a quantitative model.

A. The mutant T7 RNAP promoters were cloned to drive the expression of sfGFP (TU3) and assembled along with the other two TUs to form circuit variants – SK267-SK272. B. The output levels obtained were characterized in DH10B and the relative strengths of expression were compared to that predicted by the model.

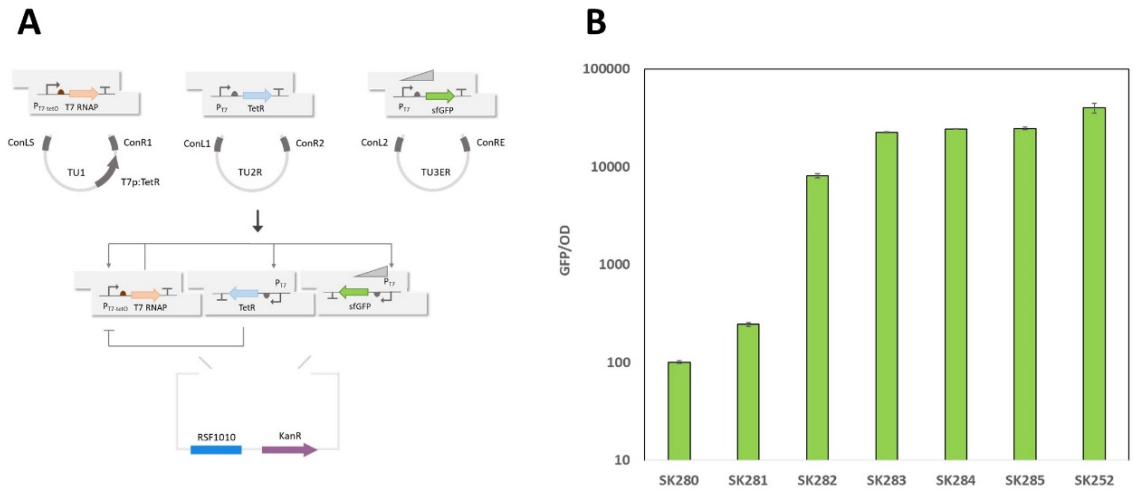

Figure 4.7: Control of output production under T7 RNAP homeostasis circuit using different RBSs driving the reporter expression.

A For the T7 RNAP homeostasis circuit (SK252), the output gene is placed under the control of RBS of varying strengths. **B** The sfGFP expression from each construct was characterized in E.coli DH10B cells.

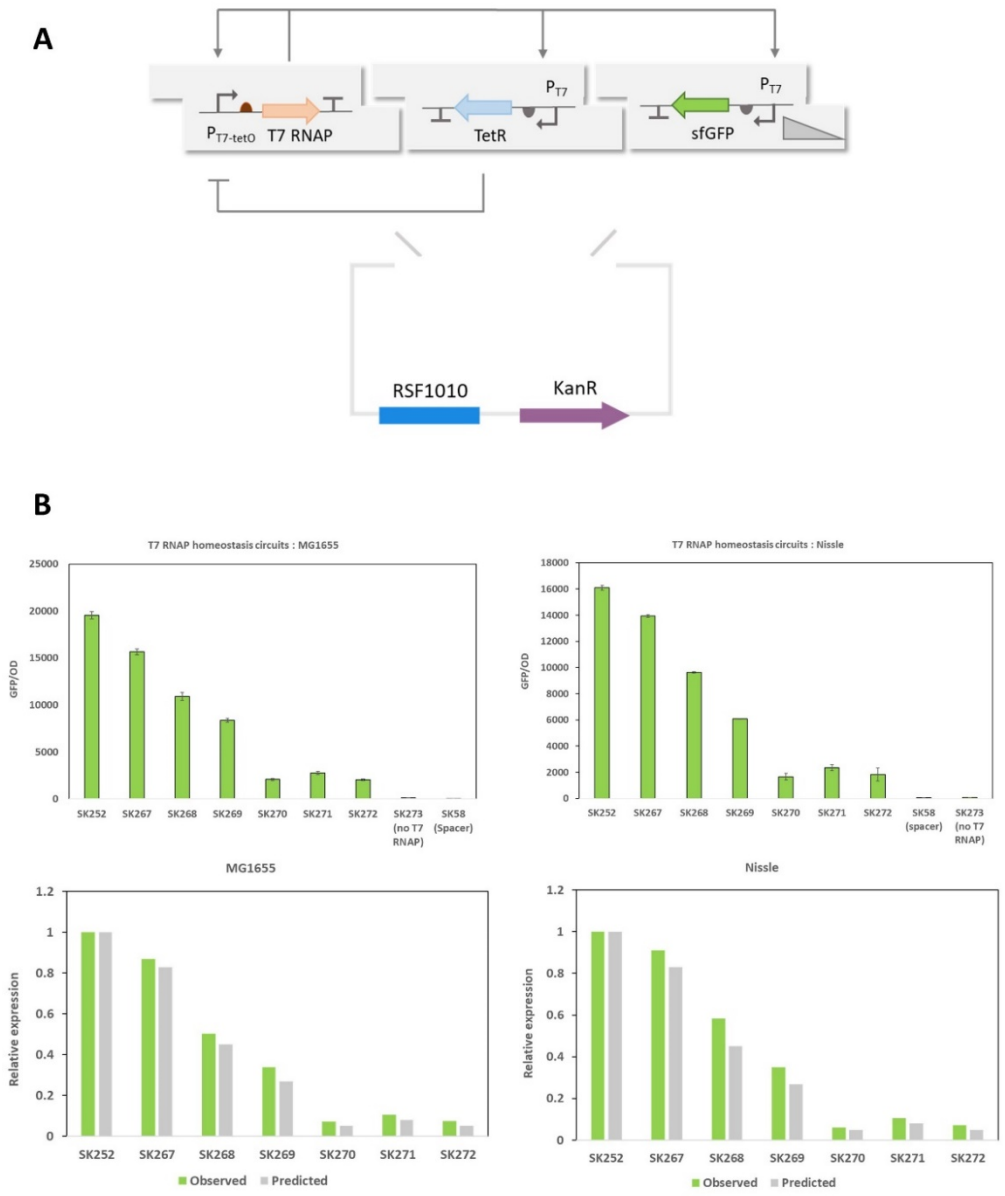

Figure 4.8: Characterization of T7 RNAP homeostasis circuits in E.coli MG1655 and Nissle strains.

A. The broad host range one plasmid encoding T7 RNAP homeostasis circuit was transformed in *E.coli* strains – MG1655 and Nissle. **B.** For each species the output expression levels (sfGFP) was characterized. The relative strengths of output levels obtained from mutant T7 RNAP promoters was determined and compared to those predicted by the model.

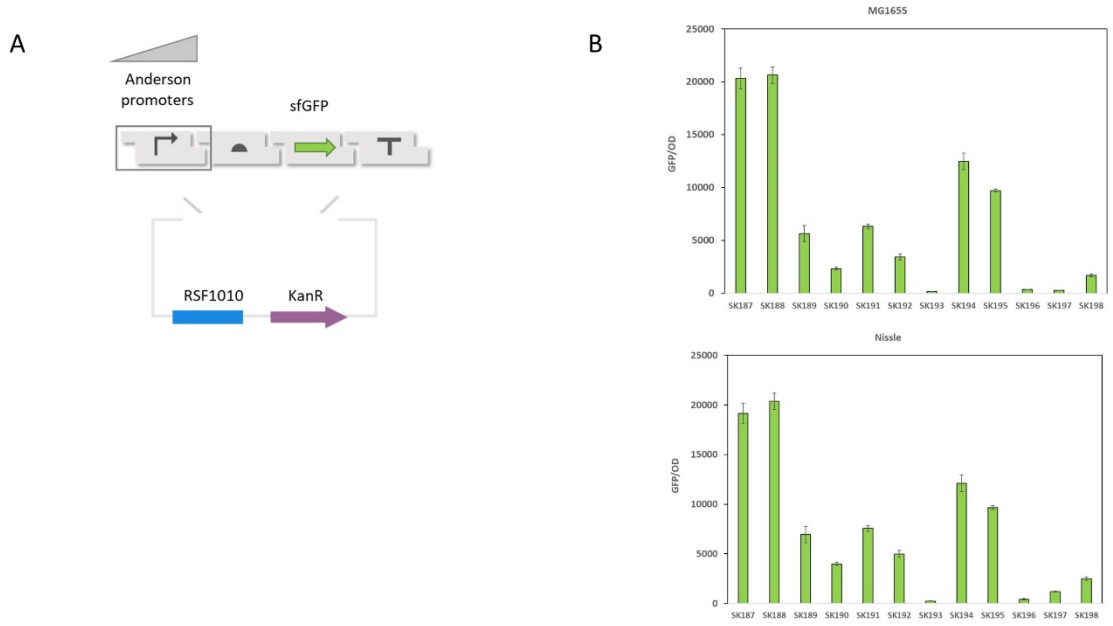

Figure 4.9: Characterization of constitutive promoters in *E.coli* MG1655 and Nissle

The panel of constitutive promoters (SK187-198) characterized in *E.coli* strains – MG1655 and Nissle. As observed from the expression of the strong constitutive promoter (SK187), the levels obtained from the T7 RNAP homeostasis circuit (SK252) was similar.

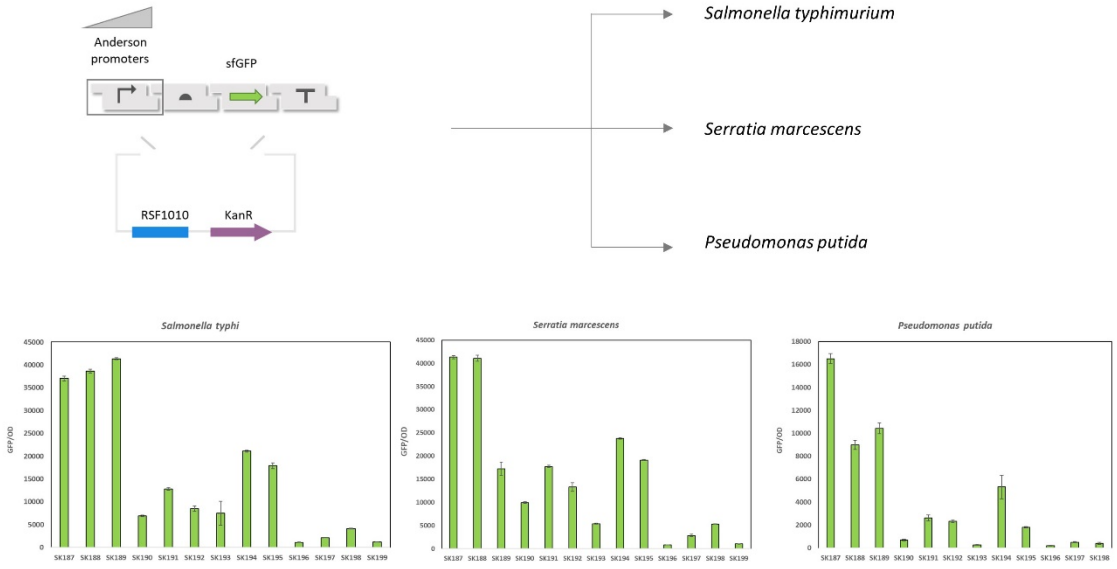

Figure 4.10 Characterization of constitutive promoters in three species: *Salmonella typhimurium, Serratia marcescens and Pseudomonas putida*

The panel of constitutive promoters (SK187-199) were characterized in the three species. The relative strengths obtained in each case is variable unlike those under T7 RNAP based regulation suggesting that orthogonal control over transcription is critical for achieving predictable behavior of genetic circuits.

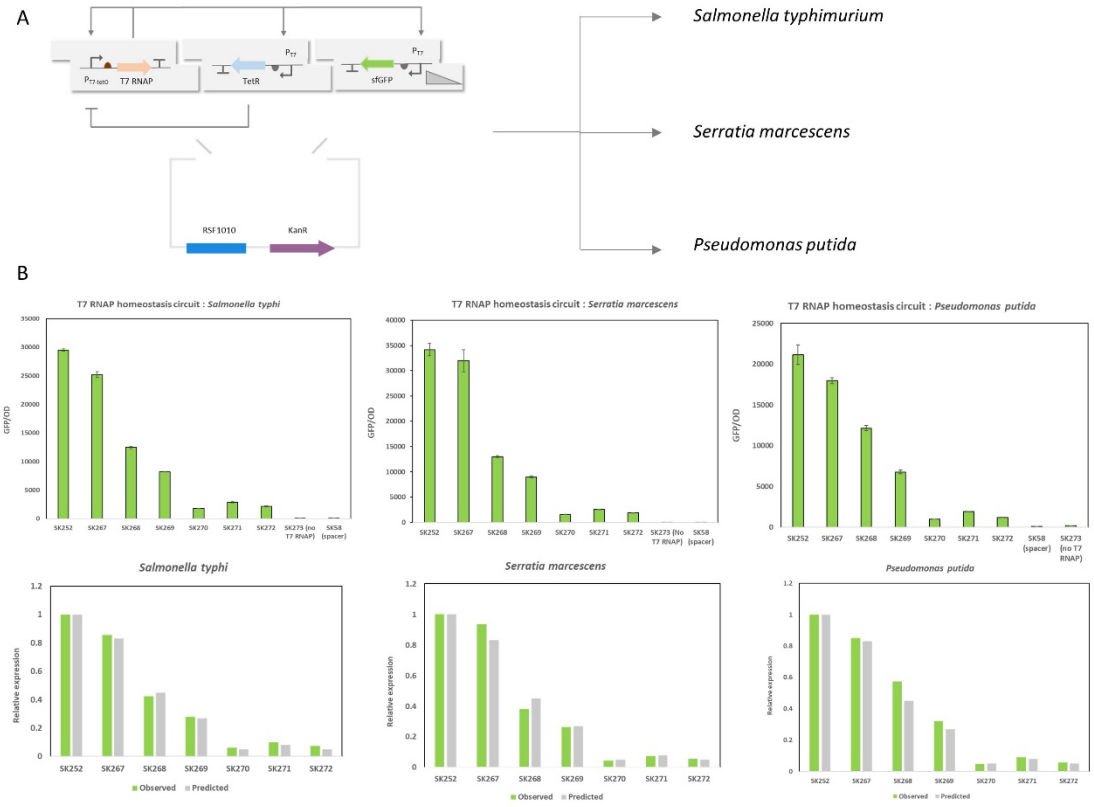

Figure 4.11: Characterization of one plasmid T7 RNAP homeostasis circuit in three species -*Salmonella typhimurium, Serratia marcescens* and *Pseudomonas putida.*

A. The broad host range one plasmid encoding T7 RNAP homeostasis circuit was transferred to three different species – *S.typhimurium, S.marcescens* and *P.putida.* **B.** The sfGFP levels obtained from the circuits in each species was characterized. The relative levels of sfGFP obtained for each mutant promoter was compared to that predicted by the model.

Table 4.1: List of genetic parts for the assembly of plasmid backbone

Name	Description	Sequence
UPS.sc101	SC101 –WT origin of replication	CCGAATGAATCCATAAAAGGCGCCTGTAGTGCCA TTTACCCCCATTCACTGCCAGAGCCGTGAGCGCA GCGAACTGAATGTCACGAAAAAGACAGCGACTC AGGTGCCTGATGGTCGGAGACAAAAGGAATATT CAGCGATTTGCCCGAGCTTGCGAGGGTGCTACTT AAGCCTTTAGGGTTTTAAGGTCTGTTTTGTAGAG GAGCAAACAGCGTTTGCGACATCCTTTTGTAATA CTGCGGAACTGACTAAAGTAGTGAGTTATACACA GGGCTGGGATCTATTCTTTTTATCTTTTTTTATTCT TTCTTTATTCTATAAATTATAACCACTTGAATATA AACAAAAAAAACACACAAAGGTCTAGCGGAATT TACAGAGGGTCTAGCAGAATTTACAAGTTTTCCA GCAAAGGTCTAGCAGAATTTACAGATACCCACAA CTCAAAGGAAAAGGACTAGTAATTATCATTGACT AGCCCATCTCAATTGGTATAGTGATTAAAATCAC CTAGACCAATTGAGATGTATGTCTGAATTAGTTG TTTTCAAAGCAAATGAACTAGCGATTAGTCGCTA TGACTTAACGGAGCATGAAACCAAGCTAATTTTA TGCTGTGTGGCACTACTCAACCCCACGATTGAAA ACCCTACAAGGAAAGAACGGACGGTATCGTTCA CTTATAACCAATACGCTCAGATGATGAACATCAG TAGGGAAAATGCTTATGGTGTATTAGCTAAAGCA ACCAGAGAGCTGATGACGAGAACTGTGGAAATC AGGAATCCTTTGGTTAAAGGCTTTAAGATTTTCC AGTGGACAAACTATGCCAAGTTCTCAAGCGAAA AATTAGAATTAGTTTTTAGTGAAGAGATATTGCC TTATCTTTTCCAGTTAAAAAAATTCATAAAATAT AATCTGGAACATGTTAAGTCTTTTGAAAACAAAT ACTCTATGAGGATTTATGAGTGGTTATTAAAAGA ACTAACACAAAAGAAAACTCACAAGGCAAATAT AGAGATTAGCCTTGATGAATTTAAGTTCATGTTA ATGCTTGAAAATAACTACCATGAGTTTAAAAGGC TTAACCAATGGGTTTTGAAACCAATAAGTAAAGA TTTAAACACTTACAGCAATATGAAATTGGTGGTT GATAAGCGAGGCCGCCCGACTGATACGTTGATTT TCCAAGTTGAACTAGATAGACAAATGGATCTCGT AACCGAACTTGAGAACAACCAGATAAAAATGAA

Table 4.1 (continued)

		TGGTGACAAAATACCAACAACCATTACATCAGAT TCCTACCTACATAACGGACTAAGAAAAACACTAC ACGATGCTTTAACTGCAAAAATTCAGCTCACCAG TTTTGAGGCAAAATTTTTGAGTGACATGCAAAGT AAGTATGATCTCAATGGTTCGTTCTCATGGCTCA CGCAAAAACAACGAACCACACTAGAGAACATAC TGGCTAAATACGGAAGGATCTGAGGTTCTTATGG CTCTTGTATCTATCAGTGAAGCATCAAGACTAAC AAACAAAAGTAGAACAACTGTTCACCGTTACATA TCAAAGGGAAAACTGTCCATATGCACAGATGAA AACGGTGTAAAAAAGATAGATACATCAGAGCTTT TACGAGTTTTTGGTGCATTCAAAGCTGTTCACCAT GAACAGATCGACAATGTAACAGATGAACAGCAT GTAACACCTAATAGAACAGGTGAAACCAGTAAA ACAAAGCAACTAGAACATGAAATTGAACACCTG AGACAACTTGTTACAGCTCAACAGTCACACATAG ACAGCCTGAAACAGGCGATGCTGCTTATCGAATC AAAGCTGCCGACAACACGGGCCCT
UPS.p15A	p15A origin of replication	CCGAGGTGCTACATTTGAAGAGATAAATTGCACT GAAATCTAGAAATATTTTATCTGATTAATAAGAT GATCTTCTTGAGATCGTTTTGGTCTGCGCGTAATC TCTTGCTCTGAAAACGAAAAAACCGCCTTGCAGG GCGGTTTTTCGAAGGTTCTCTGAGCTACCAACTCT TTGAACCGAGGTAACTGGCTTGGAGGAGCGCAGT CACCAAAACTTGTCCTTTCAGTTTAGCCTTAACCG GCGCATGACTTCAAGACTAACTCCTCTAAATCAA TTACCAGTGGCTGCTGCCAGTGGTGCTTTTGCAT GTCTTTCCGGGTTGGACTCAAGACGATAGTTACC GGATAAGGCGCAGCGGTCGGACTGAACGGGGGG TTCGTGCATACAGTCCAGCTTGGAGCGAACTGCC TACCCGGAACTGAGTGTCAGGCGTGGAATGAGA CAAACGCGGCCATAACAGCGGAATGACACCGGT AAACCGAAAGGCAGGAACAGGAGAGCGCACGAG GGAGCCGCCAGGGGGAAACGCCTGGTATCTTTAT AGTCCTGTCGGGTTTCGCCACCACTGATTTGAGC GTCAGATTTCGTGATGCTTGTCAGGGGGGCGGAG CCTATGGAAAAACGGCTTTGCCGCGGCCCTCTCA CTTCCCTGTTAAGTATCTTCCTGGCATCTTCCAGG AAATCTCCGCCCCGTTCGTAAGCCATTTCCGCTC GCCGCAGTCGAACGACCGAGCGTAGCGAGTCAG

Table 4.1 (continued)

		TGAGCGAGGAAGCGGAATATATCCTGTATCACAT ATTCTGCTGACGCACCGGTGCAGCCTTTTTTCTCC TGCCACATGAAGCACTTCACTGACACCCTCATCA GTGCCAACATAGTAAGCCAGTATACACTCCGCTA GCGCTGATGTCCGGCGGTGCTTTTGCCGTTACGC ACCACCCCGTCAGTAGCTGAACAGGCCCT
UPS.pBR322-rop	Pbr322 origin of replication containing rop protein	CCGATCCATCCTTTTTCGCACGATATACAGGATTT TGCCAAAGGGTTCGTGTAGACTTTCCTTGGTGTA TCCAACGGCGTCAGCCGGGCAGGATAGGTGAAG TAGGCCCACCCGCGAGCGGGTGTTCCTTCTTCAC TGTCCCTTATTCGCACCTGGCGGTGCTCAACGGG AATCCTGCTCTGCGAGGCTGGCCGTAGGCCGGCC CCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAA AAAAACCACCGCTACCAGCGGTGGTTTGTTTGCC GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTA ACTGGCTTCAGCAGAGCGCAGATACCAAATACTG TCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTC AAGAACTCTGTAGCACCGCCTACATACCTCGCTC TGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGG CGATAAGTCGTGTCTTACCGGGTTGGACTCAAGA CGATAGTTACCGGATAAGGCGCAGCGGTCGGGCT GAACGGGGGGTTCGTGCACACAGCCCAGCTTGG AGCGAACGACCTACACCGAACTGAGATACCTAC AGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGA AGGGAGAAAGGCGGACAGGTATCCGGTAAGCGG CAGGGTCGGAACAGGAGAGCGCACGAGGGAGCT TCCAGGGGGAAACGCCTGGTATCTTTATAGTCCT GTCGGGTTTCGCCACCTCTGACTTGAGCGTCGAT TTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATG GAAAAACGCCAGCAACGCGGCCTTTTTACGGTTC CTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTT TCCTGCGTTATCCCCTGATTCTGTGGATAACCGTA TTACCGCCTTTGAGTGAGCTGATACCGCTCGCCG CAGCCGAACGACCGAGCGCAGCGAGTCAGTGAG CGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTT CTCCTTACGCATCTGTGCGGTATTTCACACCGCAT ATGGTGCACTCTCAGTACAATCTGCTCTGATGCC GCATAGTTAAGCCAGTATACACTCCGCTATCGCT ACGTGACTGGGTCATGGCTGCGCCCCGACACCCG

Table 4.1 (continued)

		CCAACACCCGCTGACGCGCCCTGACGGGCTTGTC TGCTCCCGGCATCCGCTTACAGACAAGCTGTGAC AGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCA CCGTCATCACCGAAACGCGCGAGGCAGCTGCGGT AAAGCTCATCAGCGTGGTCGTGCAGCGATTCACA GATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGA GTTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATA AAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTT TGGTCACTGATGCCTCCGTGTAAGGGGGATTTCT GTTCATGGGGGTAATGATACCGATGAAACGAGA GAGGATGCTCACGATACGGGTTACTGATGATGAA CATGCCCGGTTACTGGAACGTTGTGAGGGTAAAC AACTGGCGGTATGGCCCAGCCCT
UPS.pUC	pUC origin of replication	CCGAACGGTTATCCACAGAATCAGGGGATAACG CAGGAAAGAACATGTGAGCAAAAGGCCAGCAAA AGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGG CGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCA TCACAAAAATCGACGCTCAAGTCAGAGGTGGCG AAACCCGACAGGACTATAAAGATACCAGGCGTTT CCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCC GACCCTGCCGCTTACCGGATACCTGTCCGCCTTTC TCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTC ACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTT CGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCG TTCAGCCCGACCGCTGCGCCTTATCCGGTAACTA TCGTCTTGAGTCCAACCCGGTAAGACACGACTTA TCGCCACTGGCAGCAGCCACTGGTAACAGGATTA GCAGAGCGAGGTATGTAGGCGGTGCTACAGAGT TCTTGAAGTGGTGGCCTAACTACGGCTACACTAG AAGAACAGTATTTGGTATCTGCGCTCTGCTGAAG CCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTT GATCCGGCAAACAAACCACCGCTGGTAGCGGTG GTTTTTTTGTTTGCAAGCAGCAGATTACGCGCAG AAAAAAAGGATCTCAAGAAGATCCTTTGATCTTT TCTACGGGGTCTGACGCTCAGTGGAACGAAAACT CACGTTAAGGGATTTTGGTCATGACCCT

Table 4.1 (continued)

UPS.RSF1030	RSF1030 origin of replication	CCGAAGGCATGCAGCGCTCTTCCGCTTCCTCGCT CACTGACTCGCTACGCTCGGTCGTTCGACTGCGG CGAGCGGTGTCAGCTCACTCAAAAGCGGTAATAC GGTTATCCACAGAATCAGGGGATAAAGCCGGAA AGAACATGTGAGCAAAAAGCAAAGCACCGGAAG AAGCCAACGCCGCAGGCGTTTTTCCATAGGCTCC GCCCCCCTGACGAGCATCACAAAAATCGACGCTC AAGCCAGAGGTGGCGAAACCCGACAGGACTATA AAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTC GTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGG ATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTG GCGCTTTCTCATAGCTCACGCTGTTGGTATCTCAG TTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGT GTGCACGAACCCCCCGTTCAGCCCGACCGCTGCG CCTTATCCGGTAACTATCGTCTTGAGTCCAACCC GGTAAGACACGACTTATCGCCACTGGCAGCAGCC ATTGGTAACTGATTTAGAGGACTTTGTCTTGAAG TTATGCACCTGTTAAGGCTAAACTGAAAGAACAG ATTTTGGTGAGTGCGGTCCTCCAACCCACTTACCT TGGTTCAAAGAGTTGGTAGCTCAGCGAACCTTGA GAAAACCACCGTTGGTAGCGGTGGTTTTTCTTTA TTTATGAGATGATGAATCAATCGGTCTATCAAGT CAACGAACAGCTATTCCGTTACTCTAGATTTCAG TGCAATTTATCTCTTCAAATGTAGCACCTGAAGT CAGCCCCATACGATATAAGTTGTAATTCTCATGT TAGTCATGCCCCGCGCCCACCGGAAGGAGCTGAC TGGGTTGAAGGCTCTCAAGGGCATCGGTCGAGAT CCCGGTGCCTAATGAGTGAGCCCT
UPS.RSF1010	RSF1010 origin of replication	CCGACGCCCTAGACAGCTGGGCGCGCCCTAACTG TCACGAACCCCTGCAATAACTGTCACGCCCCCCT GCAATAACTGTCACGAACCCCTGCAATAACTGTC ACGCCCCCAAACCTGCAAACCCAGCAGGGGCGG GGGCTGGCGGGGTGTTGGAAAAATCCATCCATGA TTATCTAAGAATAATCCACTAGGCGCGGTTATCA GCGCCCTTGTGGGGCGCTGCTGCCCTTGCCCAAT ATGCCCGGCCAGAGGCCGGATAGCTGGTCTATTC GCTGCGCTAGGCTACACACCGCCCCACCGCTGCG CGGCAGGGGGAAAGGCGGGCAAAGCCCGCTAAA CCCCACACCAAACCCCGCAGAAATACGCTGGAG CGCTTTTAGCCGCTTTAGCGGCCTTTCCCCCTACC

Table 4.1 (continued)

		CGAAGGGTGGGGGCGCGTGTGCAGCCCCGCAGG GCCTGTCTCGGTCGATCATTCAGCCCGGCTCATC CTTCGGATCTAGCTACACTTTATGCTTCCGGCTCG TATAATGTGTGGGAATCGCTTGGTGGCCCGATGA AGAACGACAGGACTTTGCAGGCCATAGGCCGAC AGCTCAAGGCCATGGGCTGTGAGCGCTTCGATAT CGGCGTCAGGGACGCCACCACCGGCCAGATGAT GAACCGGGAATGGTCAGCCGCCGAAGTGCTCCA GAACACGCCATGGCTCAAGCGGATGAATGCCCA GGGCAATGACGTGTATATCAGGCCCGCCGAGCA GGAGCGGCATGGTCTGGTGCTGGTGGACGACCTC AGCGAGTTTGACCTGGATGACATGAAAGCCGAG GGCCGGGAGCCTGCCCTGGTAGTGGAAACCAGC CCGAAGAACTATCAGGCATGGGTCAAGGTGGCC GACGCCGCAGGCGGTGAACTTCGGGGGCAGATT GCCCGGACGCTGGCCAGCGAGTACGACGCCGAC CCGGCCAGCGCCGACAGCCGCCACTATGGCCGCT TGGCGGGCTTCACCAACCGCAAGGACAAGCACA CCACCCGCGCCGGTTATCAGCCGTGGGTGCTGCT GCGTGAATCCAAGGGCAAGACCGCCACCGCTGG CCCGGCGCTGGTGCAGCAGGCTGGCCAGCAGATC GAGCAGGCCCAGCGGCAGCAGGAGAAGGCCCGC AGGCTGGCCAGCCTCGAACTGCCCGAGCGGCAG CTTAGCCGCCACCGGCGCACGGCGCTGGACGAGT ACCGCAGCGAGATGGCCGGGCTGGTCAAGCGCTT CGGTGATGACCTCAGCAAGTGCGACTTTATCGCC GCGCAGAAGCTGGCCAGCCGGGGCCGCAGTGCC GAGGAAATCGGCAAGGCCATGGCCGAGGCCAGC CCAGCGCTGGCAGAGCGCAAGCCCGGCCACGAA GCGGATTACATCGAGCGCACCGTCAGCAAGGTCA TGGGTCTGCCCAGCGTCCAGCTTGCGCGGGCCGA GCTGGCACGGGCACCGGCACCCCGCCAGCGAGG CATGGACAGGGGCGGGCCAGATTTCAGCATGTA GTGCTTGCGTTGGTACTCACGCCTGTTATACTATG AGTACTCACGCACAGAAGGGGGTTTTATGGAATA CGAAAAAAGCGCTTCAGGGTCGGTCTACCTGATC AAAAGTGACAAGGGCTATTGGTTGCCCGGTGGCT TTGGTTATACGTCAAACAAGGCCGAGGCTGGCCG CTTTTCAGTCGCTGATATGGCCAGCCTTAACCTTG ACGGCTGCACCTTGTCCTTGTTCCGCGAACACAA

Table 4.1 (continued)

		GCCTTTCGGCCCCGGCAAGTTTCTCGGTGACTGA TATGAAAGACCAAAAGGACAAGCAGACCGGCGA CCTGCTGGCCAGCCCTGACGCTGTACGCCAAGCG CGATATGCCGAGCGCATGAAGGCCAAAGGGATG CGTCAGCGCAAGTTCTGGCTGACCGACGACGAAT ACGAGGCGCTGCGCGAGTGCCTGGAAGAACTCA GAGCGGCGCAGGGCGGGGGTAGTGACCCCGCCA GCGCCTAACCACCAACTGCCTGCAAAGGAGGCA ATCAATGGCTACCCATAAGCCTATCAATATTCTG GAGGCGTTCGCAGCAGCGCCGCCACCGCTGGACT ACGTTTTGCCCAACATGGTGGCCGGTACGGTCGG GGCGCTGGTGTCGCCCGGTGGTGCCGGTAAATCC ATGCTGGCCCTGCAACTGGCCGCACAGATTGCAG GCGGGCCGGATCTGCTGGAGGTGGGCGAACTGC CCACCGGCCCGGTGATCTACCTGCCCGCCGAAGA TCCGCCCACCGCCATTCATCACCGCCTGCACGCC CTTGGGGCGCACCTCAGCGCCGAGGAACGGCAA GCCGTGGCTGACGGCCTGCTGATCCAGCCGCTGA TCGGCAGCCTGCCCAACATCATGGCCCCGGAGTG GTTCGACGGCCTCAAGCGCGCCGCCGAGGGCCGC CGCCTGATGGTGCTGGACACGCTGCGCCGGTTCC ACATCGAGGAAGAAAACGCCAGCGGCCCCATGG CCCAGGTCATCGGTCGCATGGAGGCCATAGCCGC CGATACCGGGTGCTCTATCGTGTTCCTGCACCAT GCCAGCAAGGGCGCGGCCATGATGGGCGCAGGC GACCAGCAGCAGGCCAGCCGGGGCAGCTCGGTA CTGGTCGATAACATCCGCTGGCAGTCCTACCTGT CGAGCATGACCAGCGCCGAGGCCGAGGAATGGG GTGTGGACGACGACCAGCGCCGGTTCTTCGTCCG CTTCGGTGTGAGCAAGGCCAACTATGGCGCACCG TTCGCTGATCGGTGGTTCAGGCGGCATGACGGCG GGGTGCTCAAGCCCGCCGTGCTGGAGAGGCAGC GCAAGAGCAAGGGGGTGCCCCGTGGTGAAGCCT AAGAACAAGCACAGCCTCAGCCACGTCCGGCAC GACCCGGCGCACTGTCTGGCCCCCGGCCTGTTCC GTGCCCTCAAGCGGGGCGAGCGCAAGCGCAGCA AGCTGGACGTGACGTATGACTACGGCGACGGCA AGCGGATCGAGTTCAGCGGCCCGGAGCCGCTGG GCGCTGATGATCTGCGCATCCTGCAAGGGCTGGT GGCCATGGCTGGGCCTAATGGCCTAGTGCTTGGC

Table 4.1 (continued)

		CCGGAACCCAAGACCGAAGGCGGACGGCAGCTC CGGCTGTTCCTGGAACCCAAGTGGGAGGCCGTCA CCGCTGATGCCATGGTGGTCAAAGGTAGCTATCG GGCGCTGGCAAAGGAAATCGGGGCAGAGGTCGA TAGTGGTGGGGCGCTCAAGCACATACAGGACTGC ATCGAGCGCCTTTGGAAGGTATCCATCATAGCCC AGAATGGCCGCAAGCGGCAGGGGTTTCGGCTGCT GTCGGAGTACGCCAGCGACGAGGCGGACGGGCG CCTGTACGTGGCCCTGAACCCCTTGATCGCGCAG GCCGTCATGGGTGGCGGCCAGCATGTGCGCATCA GCATGGACGAGGTGCGGGCGCTGGACAGCGAAA CCGCCCGCCTGCTGCACCAGCGGCTGTGTGGCTG GATCGACCCCGGCAAAACCGGCAAGGCTTCCATA GATACCTTGTGCGGCTATGTCTGGCCGTCAGAGG CCAGTGGTTCGACCATGCGCAAGCGCCGCCAGCG GGTGCGCGAGGCGTTGCCGGAGCTGGTCGCGCTG GGCTGGACGGTAACCGAGTTCGCGGCGGGCAAG TACGACATCACCCGGCCCAAGGCGGCAGGCTGA GGCCGGCCTACGGCCAGCCTCGCAGAGCAGGATT CCCGTTGAGCACCGCCAGGTGCGAATAAGGGAC AGTGAAGAAGGAACACCCGCTCGCGGGTGGGCC TACTTCACCTATCCTGCCCGGCTGACGCCGTTGG CCCT
UPS.CamR	CamR resistance cassette	TACACGCACCAATAAAAAACGCCCGGCGGCAAC CGAGCGTTCTGAACAAATCCAGATGGAGTTCTGA GGTCATTACTGGATCTATCAACAGGAGTCCAAGC GAGCTCGATATCAAATTACGCCCCGCCCTGCCAC TCATCGCAGTACTGTTGTAATTCATTAAGCATTCT GCCGACATGGAAGCCATCACAAACGGCATGATG AACCTGAATCGCCAGCGGCATCAGCACCTTGTCG CCTTGCGTATAATATTTGCCCATGGTGAAAACGG GGGCGAAGAAGTTGTCCATATTGGCCACGTTTAA ATCAAAACTGGTGAAACTCACCCAGGGATTGGCT GAAACGAAAAACATATTCTCAATAAACCCTTTAG GGAAATAGGCCAGGTTTTCACCGTAACACGCCAC ATCTTGCGAATATATGTGTAGAAACTGCCGGAAA TCGTCGTGGTATTCACTCCAGAGCGATGAAAACG TTTCAGTTTGCTCATGGAAAACGGTGTAACAAGG GTGAACACTATCCCATATCACCAGCTCACCGTCT TTCATTGCCATACGAAATTCCGGATGAGCATTCA

Table 4.1 (continued)

		TCAGGCGGGCAAGAATGTGAATAAAGGCCGGAT AAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAA AAGGCCGTAATATCCAGCTGAACGGTCTGGTTAT AGGTACATTGAGCAACTGACTGAAATGCCTCAAA ATGTTCTTTACGATGCCATTGGGATATATCAACG GTGGTATATCCAGTGATTTTTTTCTCCATTTTAGC TTCCTTAGCTCCTGAAAATCTCGATAACTCAAAA AATACGCCCGGTAGTGATCTTATTTCATTATGGT GAAAGTTGGAACCTCTTACGTGCCCGATCAACCT GTCCGA
UPS.KanR	KanR resistance cassette	TACAGGGGACGGGCCCTGCGGCGCGCGTCACTCG GTCTGTGCGGCGTGCCCGAGGGTAGACCAAAAA AAAAAAAACACCCGTTAGGGTGTTCAATAATTGG CCCGCTTAGAAAAACTCATCGAGCATCAAATGAA ACTGCAATTTATTCATATCAGGATTATCAATACC ATATTTTTGAAAAAGCCGTTTCTGTAATGAAGGA GAAAACTCACCGAGGCAGTTCCATAGGATGGCA AGATCCTGGTATCGGTCTGCGATTCCGACTCGTC CAACATCAATACAACCTATTAATTTCCCCTCGTC AAAAATAAGGTTATCAAGTGAGAAATCACCATG AGTGACGACTGAATCCGGTGAGAATGGCAAAAG CTTATGCATTTCTTTCCAGACTTGTTCAACAGGCC AGCCATTACGCTCGTCATCAAAATCACTCGCATC AACCAAACCGTTATTCATTCGTGATTGCGCCTGA GCGAGGCGAAATACGCGATCGCTGTTAAAAGGA CAATTACAAACAGGAATCGAATGCAACCGGCGC AGGAACACTGCCAGCGCATCAACAATATTTTCAC CTGAATCAGGATATTCTTCTAATACCTGGAATGC TGTTTTCCCGGGGATCGCAGTGGTGAGTAACCAT GCATCATCAGGAGTACGGATAAAATGCTTGATGG TCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAG TCTGACCATCTCATCTGTAACATCATTGGCAACG CTACCTTTGCCATGTTTCAGAAACAACTCTGGCG CATCGGGCTTCCCATACAATCGATAGATTGTCGC ACCTGATTGCCCGACATTATCGCGAGCCCATTTA TACCCATATAAATCAGCATCCATGTTGGAATTTA ATCGCGGCCTGGAGCAAGACGTTTCCCGTTGAAT ATGGCTCATAACACCCCTTGTATTACTGTTTATGT AAGCAGACAGTTTTATTGTTCATGATGATATATTT TTATCTTGTGCAATGTAACATCAGAGATTTTGAG

	Table 4.1 (continued)	ACACAACGTGGCTTTGTTGAATAAATCGAACTTT TGCTGAGTTGAAGGATCAGGCCGA
UPS.AmpR	AmpR resistance cassettee	TACAGATTATCAAAAAGGATCTTCACCTAGATCC TTTTAAATTAAAAATGAAGTTTTAAATCAATCTA AAGTATATATGAGTAAACTTGGTCTGACAGTTAC CAATGCTTAATCAGTGAGGCACCTATCTCAGCGA TCTGTCTATTTCGTTCATCCATAGTTGCCTGACTC CCCGTCGTGTAGATAACTACGATACGGGAGGGCT TACCATCTGGCCCCAGTGCTGCAATGATACCGCG GGACCCACGCTCACCGGCTCCAGATTTATCAGCA ATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGA AGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTC TATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGT TCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCA TTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTT GGTATGGCTTCATTCAGCTCCGGTTCCCAACGAT CAAGGCGAGTTACATGATCCCCCATGTTGTGCAA AAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTT GTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCA TGGTTATGGCAGCACTGCATAATTCTCTTACTGTC ATGCCATCCGTAAGATGCTTTTCTGTGACTGGTG AGTACTCAACCAAGTCATTCTGAGAATAGTGTAT GCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATA CGGGATAATACCGCGCCACATAGCAGAACTTTAA AAGTGCTCATCATTGGAAAACGTTCTTCGGGGCG AAAACTCTCAAGGATCTTACCGCTGTTGAGATCC AGTTCGATGTAACCCACTCGTGCACCCAACTGAT CTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGG TGAGCAAAAACAGGAAGGCAAAATGCCGCAAAA AAGGGAATAAGGGCGACACGGAAATGTTGAATA CTCATACTCTTCCTTTTTCAATATTATTGAAGCAT TTATCAGGGTTATTGTCTCATGAGCGGATACATA TTTGAATGTATTTAGAAAAATAAACAAATAGGGG TTCCGCGCACATTTCCCCGAAAAGTGCCACCTGC CGA

Table 4.2: List of elements for the construction UPS genetic circuits

Name	Description	Sequence
UPS.P25	T7p-TetO	AACGGGGAACTGCCAGACATCAAATAAAACAAA AGGCTCAGTCGGAAGACTGGGCCTTTTGTTTTAT CTGTTGTTTGTCGGTGAACACTCTCCCGATTGCGA TACAGACCCTAATTTCACATCATATGACATATGA CACTAATCTATCATTGATAGGTATAAATTAATAC GACTCACTATAGGGTGACCTATCAGTGATAGAGG GAG
UPS.P3	T7p with rnpB terminator upstream (used to drive TetR)	AACGTTATTGTTCGGTCAGTTTCACCTGATTTACG TAAAAACCCGCTTCGGCGGGTTTTTGCTTTTGGA GGGGCAGAAAGATGAATGACTGTCTCATGACCAT TCGTCTTCACCTCTAATACGACTCACTATAGGGA G
UPS.T7p4	T7p with UPS.T23 upstream (used to drive sfGFP)	AACGGAATTAGAGATCAAGCCTTAACGAACTAA GACCCCCGCACCGAAAGGTCCGGGGGTTTTTTTT GACCTTAAAAACATAACCGAGGAGCAGACATGC GGCAGGGCGGTCACACGCGGGCCCGCAGTGGCC GGCAGGCGCACGCCCTGCTCGAAGTAATACGACT CACTATAGGGAG
T7P4.215	Mutant T7 promoter used to drive sfGFP (SK267)	TAATACGACTCACTAAA
T7P4.216	Mutant T7 promoter used to drive sfGFP (SK268)	TAATACGACTCACGATA
T7P4.217	Mutant T7 promoter used to drive sfGFP (SK269)	TAATACGACTCACCATA

Table 4.2 (continued)

T7P4.218	Mutant T7 promoter used to drive sfGFP (SK270)	TAATACGACTCATTATA
T7P4.219	Mutant T7 promoter used to drive sfGFP (SK271)	TAATACGACTCCCTATA
T7P4.220	Mutant T7 promoter used to drive sfGFP (SK272)	TAATACGACTTACTATA
G4.R3	RBS used with P25 to drive T7 RNAP (SK249)	GGAGTAAGGATACATCAGGAACGGCTATG
G4.R4	RBS used with P25 to drive T7 RNAP (SK250)	GGAGTCCAATTGGGAAGACAGACTTATG
G4.R6	RBS used with P25 to drive T7 RNAP (SK252)	GGAGGCAGATAAAGGGAGCGTAGTTATG
T.R1	RBS used to drive TetR for circuits	GGAGAAATTTTATTTAATTTAAAGGAGATATAATATG

Table 4.2 (continued)

G3.R7	RBS (along with riboJ) used to drive sfGFP expression (SK252)	GGAGAGCTGTCACCGGATGTGCTTTCCGGTCTGA TGAGTCCGTGAGGACGAAACAGCCTCTACAAATA ATTTTGTTTAATAAGGAGGTATATTTATG
UPS.G4	T7 RNAP coding sequence	TATGAACACGATTAACATCGCTAAGAACGACTTC TCTGACATCGAACTGGCTGCTATCCCGTTCAACA CTCTGGCTGACCATTACGGTGAGCGTTTAGCTCG CGAACAGTTGGCCCTTGAGCATGAGTCTTACGAG ATGGGTGAAGCACGCTTCCGCAAGATGTTTGAGC GTCAACTTAAAGCTGGTGAGGTTGCGGATAACGC TGCCGCCAAGCCTCTCATCACTACCCTACTCCCTA AGATGATTGCACGCATCAACGACTGGTTTGAGGA AGTGAAAGCTAAGCGCGGCAAGCGCCCGACAGC CTTCCAGTTCCTGCAAGAAATCAAGCCGGAAGCC GTAGCGTACATCACCATTAAGACCACTCTGGCTT GCCTAACCAGTGCTGACAATACAACCGTTCAGGC TGTAGCAAGCGCAATCGGTCGGGCCATTGAGGAC GAGGCTCGCTTCGGTCGTATCCGTGACCTTGAAG CTAAGCACTTCAAGAAAAACGTTGAGGAACAAC TCAACAAGCGCGTAGGGCACGTCTACAAGAAAG CATTTATGCAAGTTGTCGAGGCTGACATGCTCTC TAAGGGTCTACTCGGTGGCGAGGCGTGGTCTTCG TGGCATAAGGAAGACTCTATTCATGTAGGAGTAC GCTGCATCGAGATGCTCATTGAGTCAACCGGAAT GGTTAGCTTACACCGCCAAAATGCTGGCGTAGTA GGTCAAGACTCTGAGACTATCGAACTCGCACCTG AATACGCTGAGGCTATCGCAACCCGTGCGGGTGC GCTGGCTGGCATCTCTCCGATGTTCCAACCTTGC GTAGTTCCTCCTAAGCCGTGGACTGGCATTACTG GTGGTGGCTATTGGGCTAACGGTCGTCGTCCTCT GGCGCTGGTGCGTACTCACAGTAAGAAAGCACTG ATGCGCTACGAAGACGTTTACATGCCTGAGGTGT ACAAAGCGATTAACATTGCGCAAAACACCGCAT GGAAAATCAACAAGAAAGTCCTAGCGGTCGCCA ACGTAATTACCAAGTGGAAGCATTGTCCGGTCGA GGACATCCCTGCGATTGAGCGTGAAGAACTCCCG ATGAAACCGGAAGACATCGACATGAATCCTGAG GCTCTGACCGCGTGGAAACGTGCTGCCGCTGCTG

Table 4.2 (continued)

		TGTACCGCAAGGACAAGGCTCGCAAGTCTCGCCG TATCAGCCTTGAGTTCATGCTTGAGCAAGCCAAT AAGTTTGCTAACCATAAGGCCATCTGGTTCCCTT ACAACATGGACTGGCGCGGTCGTGTTTACGCTGT GTCAATGTTCAACCCGCAAGGTAACGATATGACC AAAGGACTGCTTACGCTGGCGAAAGGTAAACCA ATCGGTAAGGAAGGTTACTACTGGCTGAAAATCC ACGGTGCAAACTGTGCGGGTGTCGATAAGGTTCC GTTCCCTGAGCGCATCAAGTTCATTGAGGAAAAC CACGAGAACATCATGGCTTGCGCTAAGTCTCCAC TGGAGAATACTTGGTGGGCTGAGCAAGATTCTCC GTTCTGCTTCCTTGCATTCTGCTTTGAGTACGCTG GGGTACAGCACCACGGCCTGAGCTATAACTGCTC CCTTCCGCTGGCGTTTGACGGGTCTTGCTCTGGCA TCCAGCACTTCTCCGCGATGCTCCGAGATGAGGT AGGTGGTCGCGCGGTTAACTTGCTTCCTAGTGAA ACCGTTCAGGACATCTACGGGATTGTTGCTAAGA AAGTCAACGAGATTCTACAAGCAGACGCAATCA ATGGGACCGATAACGAAGTAGTTACCGTGACCG ATGAGAACACTGGTGAAATCTCTGAGAAAGTCA AGCTGGGCACTAAGGCACTGGCTGGTCAATGGCT GGCTTACGGTGTTACTCGCAGTGTGACTAAGCGT TCAGTCATGACGCTGGCTTACGGGTCCAAAGAGT TCGGCTTCCGTCAACAAGTGCTGGAAGATACCAT TCAGCCAGCTATTGATTCCGGCAAGGGTCTGATG TTCACTCAGCCGAATCAGGCTGCTGGATACATGG CTAAGCTGATTTGGGAATCTGTGAGCGTGACGGT GGTAGCTGCGGTTGAAGCAATGAACTGGCTTAAG TCTGCTGCTAAGCTGCTGGCTGCTGAGGTCAAAG ATAAGAAGACTGGAGAGATTCTTCGCAAGCGTTG CGCTGTGCATTGGGTAACTCCTGATGGTTTCCCTG TGTGGCAGGAATACAAGAAGCCTATTCAGACGC GCTTGAACCTGATGTTCCTCGGTCAGTTCCGCTTA CAGCCTACCATTAACACCAACAAAGATAGCGAG ATTGATGCACACAAACAGGAGTCTGGTATCGCTC CTAACTTTGTACACAGCCAAGACGGTAGCCACCT TCGTAAGACTGTAGTGTGGGCACACGAGAAGTAC GGAATCGAATCTTTTGCACTGATTCACGACTCCTT CGGTACCATTCCGGCTGACGCTGCGAACCTGTTC AAAGCAGTGCGCGAAACTATGGTTGACACATATG

	Table 4.2 (continued)	AGTCTTGTGATGTACTGGCTGATTTCTACGACCA GTTCGCTGACCAGTTGCACGAGTCTCAATTGGAC AAAATGCCAGCACTTCCGGCTAAAGGTAACTTGA ACCTCCGTGACATCTTAGAGTCGGACTTCGCGTT CGCGTAAATCC
UPS.G2	TetR coding sequence	TATGTCTCGTTTAGATAAAAGTAAAGTGATTAAC AGCGCATTAGAGCTGCTTAATGAGGTCGGAATCG AAGGTTTAACAACCCGTAAACTCGCCCAGAAGCT AGGTGTAGAGCAGCCTACATTGTATTGGCATGTA AAAAATAAGCGGGCTTTGCTCGACGCCTTAGCCA TTGAGATGTTAGATAGGCACCATACTCACTTTTG CCCTTTAGAAGGGGAAAGCTGGCAAGATTTTTTA CGTAATAACGCTAAAAGTTTTAGATGTGCTTTAC TAAGTCATCGCGATGGAGCAAAAGTACATTTAGG TACACGGCCTACAGAAAAACAGTATGAAACTCTC GAAAATCAATTAGCCTTTTTATGCCAACAAGGTT TTTCACTAGAGAATGCATTATATGCACTCAGCGC AGTGGGGCATTTTACTTTAGGTTGCGTATTGGAA GATCAAGAGCATCAAGTCGCTAAAGAAGAAAGG GAAACACCTACTACTGATAGTATGCCGCCATTAT TACGACAAGCTATCGAATTATTTGATCACCAAGG TGCAGAGCCAGCCTTCTTATTCGGCCTTGAATTG ATCATATGCGGATTAGAAAAACAACTTAAATGTG AAAGTGGGTCTTAAATCC
UPS. G3	sfGFP	TATGGCATCCAAGGGCGAGGAGCTCTTTACTGGC GTAGTACCAATTCTCGTAGAGCTCGATGGCGATG TAAATGGCCATAAGTTTTCCGTACGCGGCGAGGG CGAGGGCGATGCAACTAACGGCAAGCTCACTCTC AAGTTTATTTGTACTACTGGCAAGCTCCCAGTAC CATGGCCAACTCTCGTAACTACTCTGACCTATGG CGTACAATGTTTTTCCCGCTATCCAGATCACATG AAGCAACATGATTTTTTTAAGTCCGCAATGCCAG AGGGCTATGTACAAGAGCGCACTATTAGCTTTAA GGATGATGGCACCTATAAGACTCGCGCAGAGGT AAAGTTTGAGGGCGATACTCTCGTAAATCGCATT GAGCTCAAGGGCATTGATTTTAAGGAGGATGGCA ATATTCTCGGCCATAAGCTGGAGTATAATTTCAA TTCCCATAATGTATATATTACCGCAGATAAGCAA AAGAATGGCATTAAGGCGAATTTTAAGATTCGCC ATAATGTGGAGGATGGCTCCGTACAACTCGCAGA

	Table 4.2 (continued)	TCATTATCAACAAAATACTCCAATTGGCGATGGC CCAGTACTCCTCCCAGATAATCATTATCTCTCCAC TCAATCCGTGCTCTCCAAAGATCCAAATGAGAAG CGCGATCACATGGTACTCCTGGAGTTTGTAACTG CAGCAGGCATTACTCATGGCATGGATGAGCTCTA TAAGCTCGAGCACCACCACCACCACCACTGAATC C
UPS.T1	Double T7 terminator	ATCCAGCAGCATTAGCATAACCCCTTGGGGCCTC TAAACGGGTCTTGAGGGGTTTTTTGAAAAGATCA AAGGATCTTCTGTGATAGATACTCGAACCCCTAG CCCGCTCTTATCGGGCGGCTAGGGGTTTTTTGTCC AGGTGCTACATTGCTG

Chapter 5: Construction of a modular framework for orthogonal gene expression and control in plants

In the previous chapter, I demonstrated how T7 RNAP homeostasis circuits can be utilized for the predictable expression of target gene independent of the host RNAP control in prokaryotes. While T7 RNAP has been shown to be active in eukaryotes, the lack of post-transcriptional processing of T7 RNAP derived transcripts results in almost no translation of these RNAs. Thus, such circuits can't be ported directly in eukaryotes. However, in these species orthogonal transcriptional regulation can be achieved using artificial transcription factors to regulate the activity of engineered synthetic promoters, operating independent of the host regulatory system. Specifically, in this chapter I have focused on the use of the highly programmable RNA-guided transcriptional activator – dCas9:VP64 in plants to regulate the activity of synthetic promoters, by using specific gRNAs that demonstrate minimal cross-reactivity. Interestingly, the activity of dCas9 based transcription factors has been demonstrated can be

ABSTRACT

As part of increasing the utility of synthetic biology methods for biological engineering, a wide range of toolkits have begun to be developed that allow for the modular construction of circuits in an increasingly diverse set of organisms. Here we have adapted a highly optimized parts and cloning framework to the modular, genome-wide regulatory control of plants. The kit is based in part on a highly programmable artificial transcription factor (ATF, dCas9:VP64), and orthogonal regulatory control of engineered circuits is

ensured via a new class of synthetic promoters. Guide RNA expression can in turn be controlled by either Pol III (U6) or ethylene-inducible Pol II promoters, implementing a fully sequence-based Orthogonal Control System (OCS). As a proof-of-concept for how the OCS can be implemented, we demonstrated the ratiometric expression of fluorescent proteins in single plants, where the relative ratio of expression could be controlled by ethylene addition. Overall, the implementation of the plant OCS toolkit should enable the bottom-up construction and manipulation of novel pathways and phenotypes in plants that are both buffered from cellular metabolism and able to control it in a highly programmable manner.

INTRODUCTION

The field of synthetic biology aims to revolutionize biotechnology by rationally engineering living organisms [10-14, 145]. One aspect of rational engineering is to embed biological organisms with complex information processing systems that can be used to control phenotypes [11, 12, 25, 132], often via synthetic gene circuits that can predictability regulate and tune expression of endogenous as well as transgenes [13, 23, 36, 138].

However the performance of such synthetic genetic circuits is often plagued by unwanted interactions between the circuit components and the host regulatory system, which can lead to loss of circuit function [36]. These unprogrammed interactions can be mitigated via the design and use of genetic parts that have minimal cross-talk with the host, creating orthogonal regulatory or orthogonal control systems (OCS) that can further serve as the basis for constructing complex genetic programs with predictable behaviors. In the

last two decades an increasing number of well-characterized genetic parts have been combined in circuits capable of complex dynamic behaviors, including bi-stable switches, oscillators, pulse generators, Boolean-complete logic gates [16, 17, 20, 25, 146]. While OCS and the circuits that comprise them were initially characterized in microbial hosts, more recently a significant fraction of them have been constructed and characterized in eukaryotic hosts such as yeast and mammalian cells [18, 27, 31, 32, 146]. More recently, synthetic transcriptional control elements have begun to be characterized in plants [147-149].

While a variety of synthetic plant transcription factors containing diverse DNA binding domains and plant-specific regulatory sequences are known [147, 149], orthogonal control requires more programmable DNA binding domains and modular regulatory domains [42, 69, 147, 149]. To this end, we describe an alternate strategy for the construction of orthogonal transcriptional regulatory elements in plants, powered by a single universal transcriptional factor – dCas9:VP64 which has been shown to work in a wide variety of eukaryotic species, including plants [31, 44, 45]. While this transcription factor has primarily been used for the regulation of endogenous genes [44, 45, 70], here we characterize a modular framework for the more universal design and use of synthetic promoters that rely only on the production of specific gRNAs to program dCas9:VP64, and the use of this set of mutually orthogonal promoters for the bottom-up construction of circuits that show multiplexed control of gene expression.

RESULTS

Construction of the APT (Advanced Plant Technology) toolkit

Traditionally the process of construction of these synthetic gene expression systems has relied on time-consuming practices of recombinant DNA technology like design of custom primers, PCR amplification, gel extraction of PCR products. Over the last decade the advent of high-throughput cloning techniques, such as Golden-gate cloning with Type IIS restriction enzymes, has greatly accelerated the design-build-test cycle for the construction and prototyping of synthetic circuits [25, 28, 138, 139]. Because the overlaps for assemblies can be modularly specified, multiple parts can be assembled sequentially in a single tube reaction.

While a Golden-Gate framework was previously described for the construction of plant expression vectors [150], here we used the highly optimized modular cloning (MoClo) framework, instantiated as a yeast toolkit (YTK) as the basis of our architecture [28]. Recently, beyond yeast expression vectors, this framework has been adapted for the construction of a mammalian toolkit (MTK) [138]. Along with both YTK and MTK, a plant toolkit based on the YTK architecture will prove essential for seamlessly porting parts and circuits across diverse eukaryotes. Briefly, in this framework the entire vector is divided into particular 'part' types flanked by BsaI restriction sites followed by a unique ligation site. Promoters, genes and terminators are generally categorized into Type 2, 3 and 4 parts respectively where each part type has a unique overhang that dictates the compatibility between part types [28, 138] (**Fig 5.1, 5.2**). This preserves the architecture of each transcriptional unit (promoter-gene-terminator). For the assembly of multiple

transcriptional units (TU), each transcriptional unit is first cloned into an 'intermediate' vector flanked by connector sequences that dictate the order of the TUs to be stitched together. By using appropriate connectors, each TU can be further assembled into a final expression vector in a single pot reaction (**Fig 5.2**) [20]. This modular approach enables rapid assembly of increasingly complex genetic circuits comprised of multiple transcriptional units.

Since *Agrobacterium*-based transformation has been the staple for plant genetic engineering for decades [151], we used compatible vectors as the basis for our toolkit, and designed and constructed three YTK-compatible shuttle vectors. Each expression vector contains the pVS1 replicon (an Agrobacterium origin of replication – OriV and two supporting proteins – RepA and StaA) and pBR22 origin for propagation in *Agrobacterium* and *E.coli* respectively, and a common antibiotic selection cassette (KanR) that has been shown to be functional in both species (**Fig 5.1, Materials and Methods**) [139, 150]. The three constructs otherwise differed in the plant selection marker - BASTA, hygromycin, and kanamycin. The resistance markers were expressed from the Nos promoter and also contained a Nos terminator [150] (**Fig 5.1**). The backbone also contains a GFP drop-out cassette that allows easy identification of correct assemblies, which should appear as colonies that lack fluorescence [28, 138](**Fig 5.1**).

Assaying reporter expression using the APT toolkit

Fluorescence and luminescence reporters are frequently used to study protein localization and interaction in plants and animals [152]. To provide these useful reporter parts in the context of our system, we cloned the strong promoter from Cauliflower mosaic virus (35S) as a Type 2 part and its corresponding terminator as a Type 4 part [153, 154]. These parts can be matched with a number of fluorescent reporter genes (GFP, BFP, YFP and RFP) all as Type 3 parts for robust reporter expression. Combinations of these proteins can also potentially be used for BIFC (Bimolecular Fluorescence Complementation) [155]. Similarly, luciferase is commonly used in plant molecular biology to study circadian rhythm [156], test the spatiotemporal activities of regulatory elements [157], and to study the plant immune system [158, 159]. The toolkit therefore also contains a luciferase gene from *Photinus pyralis,* commonly known as firefly luciferase (F-luc) [148].

Single TUs comprised of a 35S promoter, fluorescent reporter genes and the luciferase gene, and a terminator that serves as a polyadenylation signal were assembled into the *Agrobacterium* shuttle expression vector (**Fig 5.3**). The activity of constructs was assayed using transient expression in *Nicotiana benthamiana* [150]. As expected, we see strong activity of the promoter with the expression of the respective reporter genes (**Fig 5.3**). In order to diversify the promoters used in circuits (and thereby avoid recombination and potentially silencing), we also included a well-characterized promoter from the Ti plasmid that drives mannopine synthase (Pmas) [160-163]. When the 35S promoter was swapped with Pmas, similar expression levels of YFP were achieved (**Fig 5.3**).

Development of an Orthogonal Control System (OCS) to regulate transgene expression

One of the primary difficulties with using synthetic biology principles and methods to engineer organisms, especially in eukaryotes, is that the functionality of synthetic circuits is often plagued by unwanted interactions of the circuit 'parts' with the underlying regulatory machinery of the host [164]. As a particularly relevant example, systems developed in the past for transgene expression caused severe growth and developmental defects in *Arabidopsis* and *Nicotiana benthamiana* [165, 166]. Therefore, it is paramount to develop regulatory tools to control transgene expression that minimizes the impact on endogenous plant machinery/physiology, while maintaining the modularity and scalability of synthetic approaches in general.

A potential solution to this problem is to develop orthogonal 'parts' that of necessity function independently of endogenous regulation by the host. For our plant toolkit, we set out to develop a fully integrated Orthogonal Control System (OCS) based on orthogonal synthetic promoters driven by an Artificial Transcription Factor (ATF). We started with the deactivated form of the Cas9 protein (dCas9) fused to the transcriptional activator domain VP64 as a highly programmable ATF [45, 70]. While dCas9:VP64 has previously been shown to upregulate the expression of endogenous genes via specific guide RNAs (gRNAs) that target the promoter region upstream of those genes [44, 167], this strategy has not been utilized for the construction of a fully orthogonal system in which custom promoters can be similarly regulated. Here we develop a suite of synthetic promoters (pATFs, promoter for Artificial Transcription Factor) in which each promoter has a similar modular architecture: varying number of repeats of gRNA binding sites followed by a

minimal 35S promoter [153, 154]. This system is inherently scalable, since new binding sites bound by new gRNAs can be built at will. The complete list of parts (promoters, genes and terminators) is provided in **Table 5.1-5.5**.

We initially varied the number of gRNA binding sites (3 and 4) upstream of the minimal 35S promoter, and analyzed expression of the reporter using transient assay in *Nicotiana benthamiana*. Three repeats provided the best expression of the reporter gene without significant background (**Fig 5.4**). The promoter architecture was further assayed for leaky expression by generating pATF:YFP/BFP/RFP constructs and expressing gRNA constitutively in the absence of dCas9:VP64 (**Fig 5.4**). None of these constructs show expression above background (**Fig 5.4**). However, upon the addition of constitutively expressed dCas9:VP64 cassette to the circuit, induction of reporter protein expression was observed (**Fig 5.4**). Each pATF demonstrated comparable levels of expression (pATF1:YFP - 3-fold, pATF3:BFP - 6-fold and pATF4:RFP - 2 fold) compared to that obtained from the regular 35S promoter (6-fold; **Fig 5.3**). The basic features of the pATF and corresponding gRNAs can thus form the basis for the OCS and should allow us to predictably control reporter and other gene circuits. The complete list of assembled OCS circuits is provided in **Table 5.6**; as the reader will see, OCS circuitry can be organized in terms of increasing complexity and demonstrates how the Design-Built-Test approach can be used to empirically generate ever more substantive plant phenotypes.

In order to show that the OCS designs could also function in stable transgenic *Arabidopsis thaliana* lines, we assembled the OCS 1-1 and 4-1 circuits **Table 5.6**; constitutive YFP and luciferase expression, respectively) in an *Agrobacterium* expression

vector containing with a kanamycin selectable marker as described previously. These OCS constructs were successfully transformed into *Arabidopsis thaliana* plants (**Fig 5.5**). As expected, the OCS 1-1 T_1 plants exhibited constitutive YFP expression (**Fig 5.5**) while the OCS 4-1 plants were imaged (as described in **Methods**) and the constitutive expression of luciferase was confirmed (**Fig 5.5**). Thus, the modular circuits assembled using the toolkit function in two species, as infiltrates in *Nicotiana* and as transgenics in *Arabidopsis.*

Inducible gene expression system via the OCS framework

The ability to precisely regulate the activity of the transgenes/circuit components based on specific input stimuli is a key feature in programmable synthetic circuits [168, 169]. In order to enable orthogonal control of induction, we designed gRNA expression cassettes to produce functional gRNAs from inducible Pol II promoters. To prevent nuclear export of gRNAs due to capping and polyadenylation, we used the hammerhead ribozyme (HHR) and Hepatitis Delta Virus (HDV) to cleave the 5' and the 3' ends of the gRNA, respectively. This strategy has been previously shown to lead to the expression of functional gRNAs from Pol II promoters, with activity similar to those driven by the Pol III (U6) promoter [170].

To proof the ribozyme processed gRNA constructs, OCS circuits were assembled where gRNAs were either expressed from a U6 promoter (OCS 1-1) or the 35S promoter (OCS 1-5), and could subsequently activate the transcription and expression of reporter genes (YFP) (**Fig 5.6**). For both OCS circuits, downstream reporter gene expression was observed, at similar levels (**Fig. 5.6**). The specific levels of gRNA obtained in each case

were analyzed using qRT-PCR (**Fig 5.6**), and as expected the level of gRNA from the strong Pol II (35S) driven expression was higher than those obtained with the U6 promoter while similar levels of reporter expression were observed for both cases, thus demonstrating that this Pol II driven gRNA expression strategy can be effectively used for OCS activation (**Fig 5.6**). For both these constructs the expression of hdCas9 (human codon optimized dCas9) was also confirmed via Western blot analysis (**Fig 5.7**).

In order to demonstrate that the Pol II-driven gRNAs could be used as part of an inducible OCS we used the well-characterized synthetic EBS promoter containing the EIN3 binding [171], and placed YFP under the downstream control of the ATF (via pATF-1) (**Fig 5.8**). This circuit (OCS1-9) should be inducible by the volatile organic compound (VOC) ethylene, which is produced from its precursor ACC (1-aminocyclopropane-1-carboxylic acid). Time-dependent expression of YFP is observed in response to 10uM ACC induction (**Fig 5.8**). Both the gRNA-1 and YFP expression levels were quantified before and after induction by qRT-PCR, a maximum of 3-fold induction was observed for both cases (**Fig 5.8**). Thus, this demonstrates that the activity from synthetic promoters can be controlled via the selective expression of the corresponding gRNAs.

Construction of a panel of mutually orthogonal synthetic promoters

Lack of multiplexed control of transgenes has been a major factor limiting the development of synthetic circuits in plants [14, 145]. Multiplexed regulation in turn requires a panel of mutually orthogonal promoters and control elements that can operate simultaneously [14, 145]. Our strategy for synthetic promoter design naturally leads to the

generation of expression cassettes that are not only orthogonal to the host but are also mutually orthogonal. The degree of orthogonality can be tuned at will via the sequence design of the multiple gRNA components. By simply minimizing homology between gRNAs, we constructed two additional promoters similar to the architecture of pATF-1, in which gRNA binding sites were followed by a minimal 35S promoter (pATF-3 and pATF-4). The orthogonality of these promoters was assayed by assembling expression constructs in which each synthetic promoter controlled the production of a unique fluorescent reporter (pATF-1: YFP, pATF-3: RFP and pATF-4: BFP). The respective gRNAs (gRNA-1, gRNA-3 and gRNA-4) were separately transcribed from a U6 promoter (**Fig 5.9**). When expression constructs were infiltrated into *Nicotiana benthamiana*, each of the synthetic promoters was specifically upregulated only when its corresponding gRNA was expressed; no background was detected from the remaining two synthetic promoters. (**Fig 5.9**).

Construction of complex ratiometric circuits

Now that we have a suite of mutually orthogonal promoters, we sought to construct simple circuits where the activity of each promoter could be independently controlled. Three separate reporter proteins were used to simultaneously monitor the activity of two promoters: pATF-1 with YFP, while both RFP and BFP were under the control of the pATF-3. By leveraging the designed, orthogonal behavior of these promoters it proved possible to construct a ratiometric circuit wherein the activity of pATF-1, and hence YFP expression, was under the control of ethylene (via ACC), while pATF-3 constitutively drove the expression of RFP and BFP (**Fig 5.10**). As expected, the addition of 10uM ACC,

induced the expression of YFP from the pATF-1 promoter (3-fold), while the expression of the other reporters remained constant (**Fig 5.10)**. The ratiometric response was further validated by qRT-PCR; pATF-1 was induced 3-fold following a similar increase in expression of gRNA-1 while there were no changes observed in the transcription of the other two reporter genes (**Fig 5.10).** The predictable behavior of the designed, artificial control elements in the ratiometric circuit is one of the first examples of complex circuitry to be described in plants, and demonstrates uniquely how natural metabolism and regulatory circuitry can be interfaced with free-standing orthogonal control systems.

DISCUSSION

While a number of modular cloning frameworks have been demonstrated in plants, there is still a lack of well-characterized orthogonal genetic parts. In particular, transcriptional orthogonality is one of the bedrocks for circuit construction in synthetic biology, and generally serves as the basis for the bottom-up construction of complex circuitry for predictable dynamics [25, 27, 36]. For eukaryotes the construction of multiple promoter elements is hindered by the typically complex regulatory sequences that lie upstream and within promoters [172-174].

The design of synthetic eukaryotic promoters has traditionally implemented a common architecture, where a strong transcriptional initiation region is cloned downstream of orthogonal DNA binding operator sequences and the latter serve as landing pads for synthetic transcription factors [42]. The engineered transcription factors have typically consisted of DNA binding proteins (i.e., prokaryotic DNA binding proteins like TetR, LacI,

LexA and PhIF [6, 68, 175]) fused to well characterized transcriptional activation domain like VP64. With the advent of programmable DNA binding proteins like zinc finger proteins and TALEs the repertoire of synthetic promoters greatly increased [42, 69, 176, 177]. That said, each new synthetic promoter still requires the construction and characterization of its own unique transcription factor [42, 177, 178].

These bottlenecks can be circumvented by the use of the highly programmable RNA-guided DNA binding protein dCas9 [45]. The dCas9 RNP fused to transcription activation domains such as VP64 has been used for the upregulation of endogenous genes in a wide variety of eukaryotic species like yeast, mammalian cells and plants [31, 44, 45, 179]. Here, we have used adapted this 'universal' transcription factor to control the expression of synthetic and orthogonal promoters without the need of addition of any other factors. Using our modular framework, we were able to quickly design and characterize a panel of mutually orthogonal promoters that could drive the production of a variety of outputs, singly and in parallel, including different fluorescent proteins (GFP, BFP, RFP and YFP) and luciferase.

The activities of dCas9 based transcription factors can potentially be controlled by simply regulating the expression of their corresponding gRNAs [27, 31], enabling the coupling of natural and synthetic transcription units, and thus natural and overlaid metabolic responses. Here we have effectively used this strategy to couple ethylene sensing (via known EIN3 binding sites) to synthetic (pATF) promoters. Moreover, by changing the number and arrangement of gRNA binding sites synthetic promoters with different levels of activation can be generated, providing further opportunities for design [180]. While it has

been previously shown that a panel of minimal plant promoters can be used with natural DNA binding sequences for modulating promoter strengths [147], the addition of completely artificial, synthetic promoters as control elements should create opportunities for increasing the specificity and strengths of engineered promoters.

Since our strategy for designing synthetic promoters is generalizable it is likely that even more complex circuits can be built by simply incorporating other transcription factor binding sites, or by changing the regulatory 'headpiece' on the dCas9 element (for example, to a repressor), [181-183].

The stabilities of genetic circuitry in plants can be greatly modified by silencing and recombination, amongst other mechanisms [160, 161, 163]. In this regard, the artificial promoter elements that we generate can potentially be crafted to avoid repetition [147], and thus to better avoid silencing and recombination. As viable artificial promoter sequences continue to accumulate, they can be compared and contrasted to identify those that are least vulnerable to modification over time. The facile addition of new parts to the standardize toolkit architecture, particularly terminators, will further increase opportunities to avoid repetition in ways that again go well beyond what is possible by relying on just a few well-characterized endogenous elements alone.

The implementation of orthogonal control systems in plants can be used to limit cross-talk between natural and overlaid regulatory elements, allowing more precise response to a variety of inputs, from VOCs to hormones to temperature, water, and nutrients. The use of orthogonal control systems to enable more precise responses to pathogenesis is especially intriguing given the presence of R genes that are specifically

responsive to individual pathogens (effector triggered immunity, ETI) [184]. The architecture we have developed is fully generalizable, and can potentially be expanded to non-model plants and other eukaryotic species such as yeast and mammalian cells by the use of appropriate transcription initiation regions under the control of similar gRNA sequences binding sites [71].

MATERIALS AND METHODS

Plasmid design and construction

The plant expression vector was generated using the plasmid pICH86966 (Addgene#48075) as the backbone. The lacZ expression cassette was replaced with the GFP dropout sequence **Table 5.2** to make the plasmid compatible with YTK architecture design. All parts described in **Table 5.1**, were cloned into the backbone pYTK001 (Addgene #65108). For the individual transcriptional units, the backbone used was pYTK095 (Addgene #65202) along with the appropriate connector sequences described in **Table 5.3**. For the design of orthogonal gRNAs, random 20-mers were generated that had a GC content of ~50%, and that were at least 5 nucleotides away from all sequences in the *Nicotiana* and *Arabidopsis* genomes. All oligonucleotides and gblocks were obtained from Integrated DNA Technologies (IDT) unless otherwise stated.

For the construction of each genetic element namely promoters, coding sequences and terminators, first they were checked for restriction sites for the following enzymes – BsmBI, BsaI, NotI and DraIII. The restriction sites in the coding sequences were removed by the use of synonymous codons while the other elements did not contain any of these

restriction sites. The complete list of parts and constructs are provided in **Table 5.1-5.6**. The part plasmids were cloned into a common vector where each genetic element is flanked by Bsa1 restriction sites followed by appropriate overhangs **Table 5.1**. For the assembly of both single TU or multi-TU, the following procedure was used: 10 fmol of backbone plasmid and 20 fmol of parts/TUs were used in a 10uL reaction with 1ul of 10x T4 ligase buffer along with 100 units of BsaI-v2 (single TU) or Esp3I (multi-TU or parts) and 100 units of T7 DNA ligase. The cycling protocol used is: 24 cycles of 3 min at 37°C (for digestion) and 5 min at 16°C (for ligation) followed by a final digestion step at 37°C for 30min and the enzymes were heat inactivated 80°C for 20 min. All constructs were transformed into DH10B cells, grown at 37°C using standard chemical transformation procedures. The colonies that lack fluorescence were inoculated and plasmids were extracted using Qiagen Miniprep kit according to the manufacturer's instructions Plasmids were maintained as the following antibiotics kanamycin (50ug/mL), chloramphenicol (34ug/mL) and carbenicillin (100ug/mL) wherever required. The plasmids were sequence verified by Sanger sequencing (UT Austin Genomic Sequencing and Analysis Facility). The correct constructs were then transformed into *Agrobacterium tumefaciens* strain GV3101 (resistant to Gentamycin and Rifampicin) and used either for transient expression in *Nicotiana benthamiana* or to generate stable lines in *Arabidopsis thaliana.* The following enzymes were used for the assemblies – BsaI-v2 (NEB #R3733S), Esp3I (NEB #R0734S) and T7 DNA ligase (NEB #M0318S).

Plant material, bacterial infiltration

Nicotiana benthamiana and *Arabidopsis thaliana* plants were grown in soil at 22°C, and 16 hr light period. For transient expression, three weeks old plants were syringe-infiltrated with *Agrobacterium tumefaciens* strain GV3101 (OD_{600} = 0.5) and leaves were imaged under Olympus BX53 Digital Fluorescence Microscope or harvested for RNA and/or protein analysis. To create stable transformation in Arabidopsis, floral dip method [185] was used. T_1 plants were selected on half MS Kanamycin (50μg/ml) plates and the selected T1 plants were analyzed using an Olympus BX53 Digital Fluorescence Microscope and a NightOwl imager for YFP expression and luciferase expression, respectively. For circuits that constitutively expressed YFP (OCS1-1) and luciferase (OCS4-1) no other obvious phenotypic differences were observed across numerous individual plants.

RNA extraction and qRT-PCR

RNA was extracted using TRIzol reagent (Ambion). 1μg total RNA was used to synthesize cDNA. After DNaseI treatment to remove any DNA contamination, random primer mix (NEB #S1330S) and M-MLV Reverse transcriptase (Invitrogen #28025-013) were used for first strand synthesis. qRT-PCR was used to quantify the RNA prepared from transient expression experiments. AzuraQuant qPCR Master Mix (Azura Genomics) was used with initial incubation at 95 °C for 2 min followed by 40 cycles of 95 °C for 10 sec and 60 °C for 30sec. Level of target RNA was calculated from the difference of threshold cycle (Ct) values between reference (*5S rRNA*) and target gene using at least three independent replicates

ACC treatment

To check the induction of reporter in response to ACC in the plasmids containing pEBS::YFP/RFP/BFP, *Nicotiana benthamiana* leaves were infiltrated with Agrobacterium; after three days post infiltration, leaf discs were cut using cork borer and incubated in either 0μM or 10μM ACC for four hours. Fluorescence microscopy was used to check YFP expression after induction.

Fluorescence and Luminescence imaging

Fluorescence microscope images after *Agrobacterium* mediated transient expression of YFP, BFP, RFP and GFP in *Nicotiana benthamiana* leaves were taken using an Olympus BX53 Digital Fluorescence Microscope. For this purpose, leaf discs were cut using cork borer from the area which was infiltrated. Images were taken using either 10X objective lens using the default filters for YFP (500/535nm), BFP (385/448nm), and RFP (560/630nm). The UV filter (350/460nm) was used to take GFP images. The exposure and gain setting were kept constant for each filter within each experiment to compare multiple leaf discs (3 to 6). In all the experiments a leaf disc from a leaf which was not infiltrated with Agrobacterium was used as a negative control in order to account for background fluorescence. All experiments were performed at least three times independently as indicated in the Results.

Expression of luciferase was detected using NightOwl II LB 983 *in vivo* imaging system (https://www.berthold.com/en/bioanalytic/products/in-vivo-imaging-systems/nightowl-lb983/). Leaves/plants were sprayed with 100μM D-luciferin, Potassium salt (GoldBio

#LUCK-300). After 5 min of incubation, images were taken in the NightOwl II LB 983. Images were captured with a backlit NightOWL LB 983 NC 100 CCD camera. Photons emitted from luciferase were collected and integrated for a 2 min period. A pseudocolor luminescent image from blue (least intense) to red (most intense), representing the distribution of the detected photons emitted from active luciferase was generated using Indigo software (Berthold Technologies).

Western blot

Total protein was extracted using urea-based denaturing buffer (100 mM NaH2PO4, 8 M urea, and 10 mM Tris-HCl, pH 8.0) and used for immunoblot analysis to check the expression. The proteins were fractionated by 8% SDS-PAGE gel and transferred to a polyvinylidene difluoride (PVDF) membrane using a transfer apparatus according to the manufacturer's protocols (Bio-Rad). The membrane was treated with 5% nonfat milk in PBS-T for 10 min for blocking, and then incubated with Cas9 antibody (Santa cruz, 7A9-3A3, 1:500) at 4 °C for overnight. After incubation, the membrane was washed three times for 5 min and incubated with horseradish peroxidase-conjugated anti-mouse (1:10000) for 2 h. The Blot was washed with PBS-T three times and detected with the ECL system (Thermo scientific, lot# SE251206).

Acknowledgements

This work was done in collaboration with the Sung lab (UT Austin).

Shaunak Kar (SK), Sibum Sung (SS) and Andrew Ellington (AE) conceived of the project. SK designed the framework and the basic elements of OCS with input from Elizabeth

Gardener, Jimmy Gollihar and Sibum Sung. SK and Yogendra Bordiya assembled all constructs. YB, Nestor Rodriguez and Junghyun Kim performed all the testing in Nicotiana with input from SS. All authors contributed with the preparation of figures. SK, YB, JK, SS and AE wrote the manuscript with input from all authors

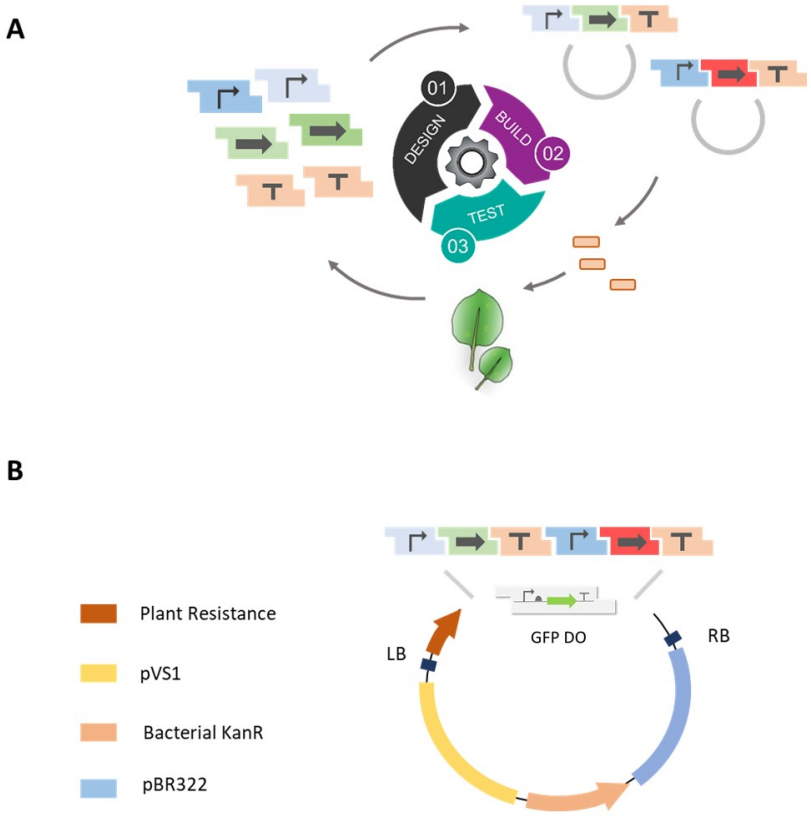

Figure 5.1. Schematic overview of the APT design-build-test cycle

A. Genetic elements such as promoters, genes and terminators are encoded as modular parts consisting of BsaI recognition sites flanked by specific overhangs to ensure the hierarchical assembly of transcriptional units. Once assembled, the constructs are transformed into Agrobacterium and the reporter expression is characterized in *Nicotiana benthamiana* leaf infiltrates **B.** Design of the shuttle vector backbone used for the assembly of constructs and subsequent propagation in *Agrobacterium*.

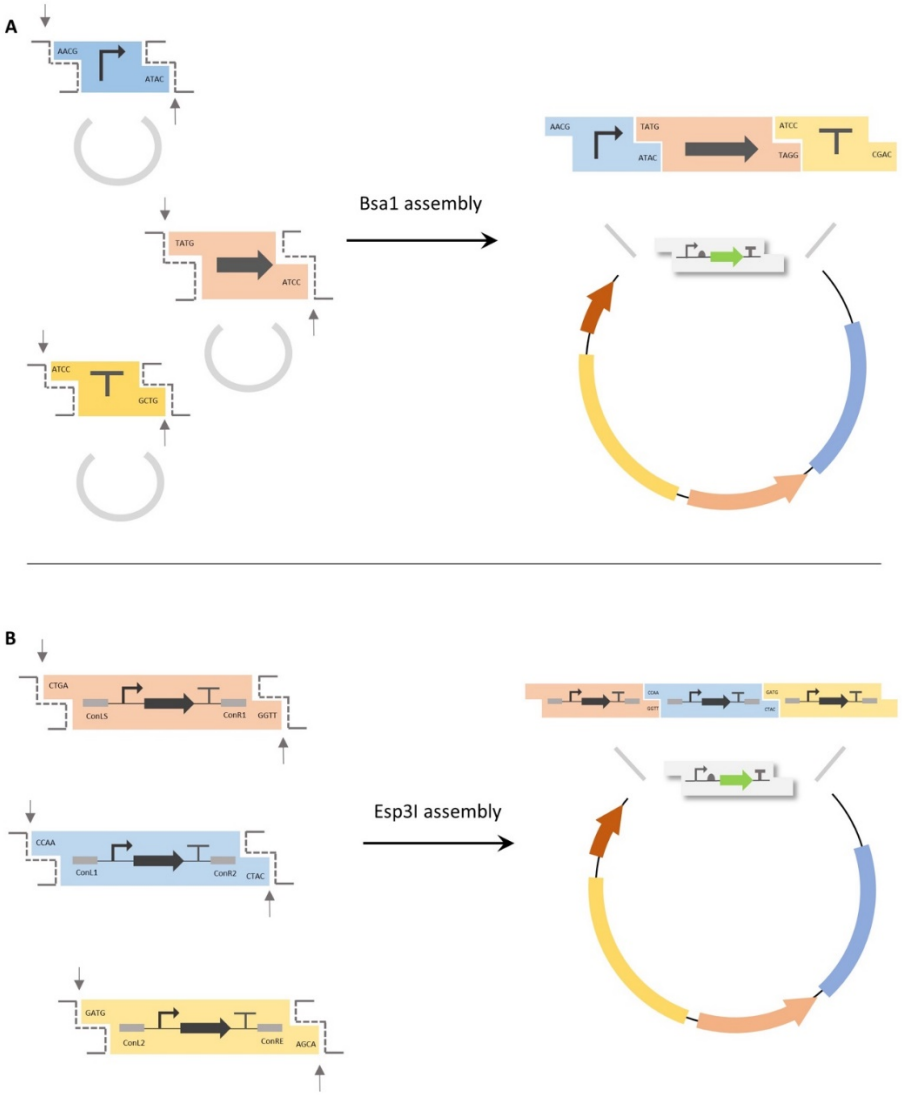

Figure 5.2: Schematic demonstrating the assembly of single (A) or multiple (B) transcriptional units into a plant expression vector.

A single transcription unit consists of a promoter, gene and a terminator parts while for multiple transcriptional units, each TU is flanked by appropriate connector sequences. The arrows depict the restriction sites for Bsa1 (A) and Esp3I (B). B) Schematic showing the assembly of multiple TUs into plant expression vector where each TU is encoded in separate plasmids flanked by appropriate connector sequences.

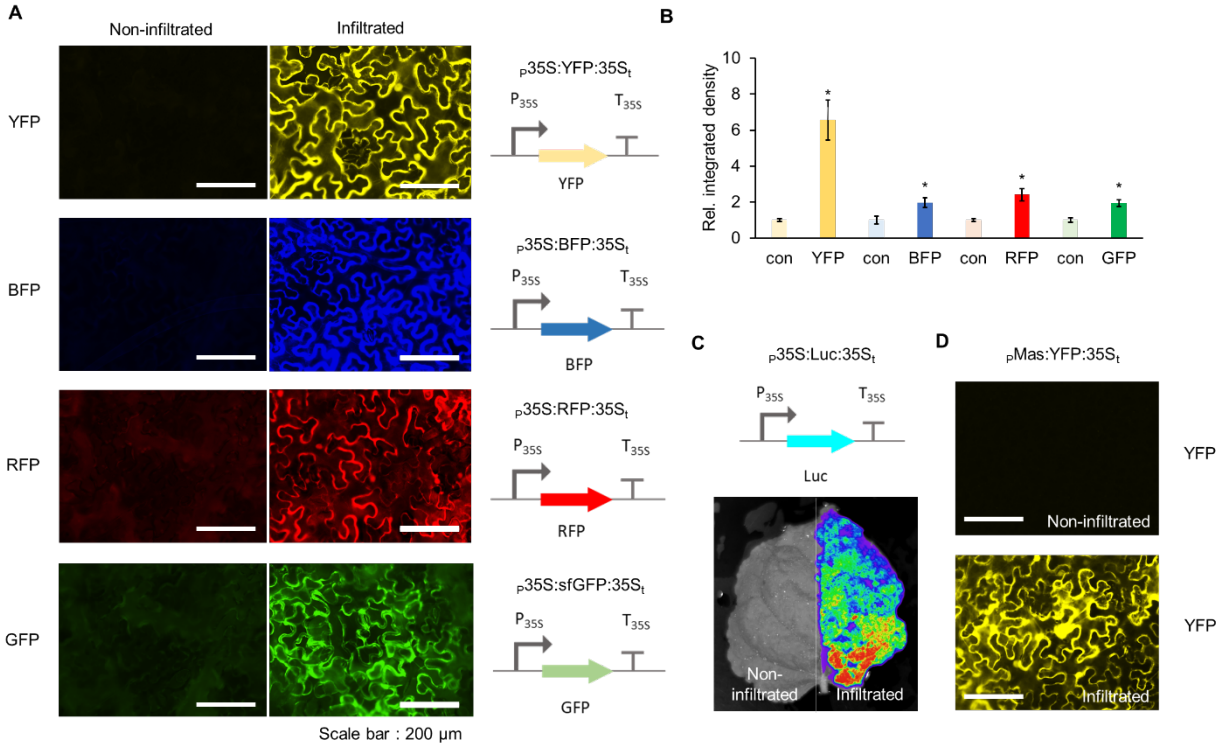

Figure 5.3. Characterization of reporter constructs assembled using APT toolkit.

A. Fluorescence microscope images showing *Agrobacterium* mediated transient expression of YFP, BFP, RFP and GFP under the control of 35S promoter into *Nicotiana benthamiana* leaves. Images on the left are from non-infiltrated leaves (negative control) captured using the appropriate filter at same exposure and gain settings as was used for the positive images on the right (**Material and Methods**). **B.** Relative integrated density of each fluorescence signal (shown in panel A). Integrated density was measured using image J software and normalized to that of a non-infiltrated control (con). Error bars: S.D. (n=3, independent replicates). Asterisks indicate statistical significance in a student t-test ($P<0.05$). **C**. Luminescence reporter luciferase expression shown by *Agrobacterium* mediated transient expression of luciferase in *Nicotiana benthamiana* leaves. Left half of the leaf was not infiltrated with *Agrobacterium*. **D.** Fluorescence microscope images showing *Agrobacterium* mediated transient expression of YFP under MAS promoter in *Nicotiana benthamiana* leaves. Image on the left is the brightfield image for the same construct.

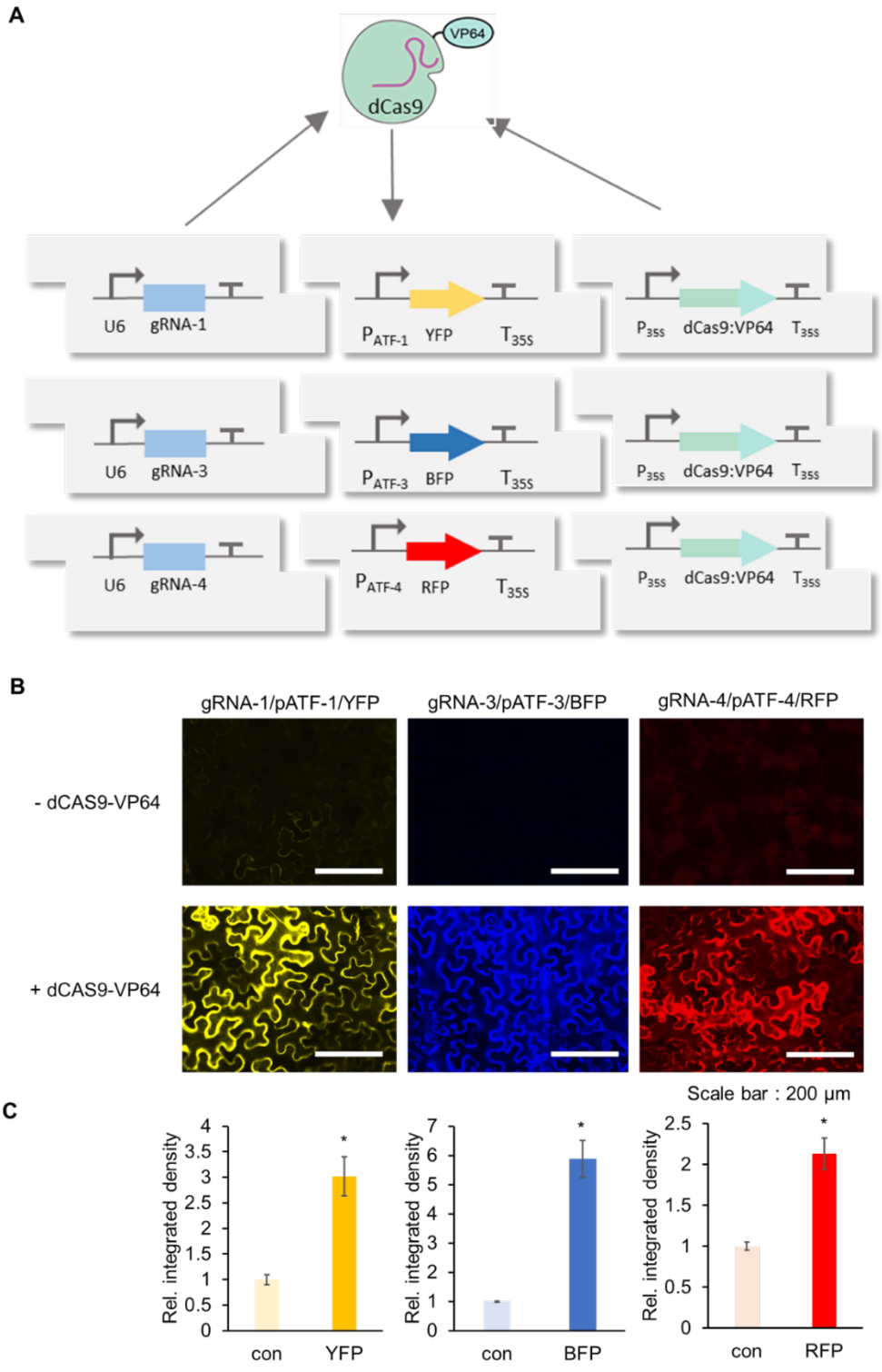
A
VP64
dCas9
U6 gRNA-1
P ATF-1 YFP T 35S
P 35S dCas9:VP64 T 35S
U6 gRNA-3
P ATF-3 BFP T 35S
P 35S dCas9:VP64 T 35S
U6 gRNA-4
P ATF-4 RFP T 35S
P 35S dCas9:VP64 T 35S
B
gRNA-1/pATF-1/YFP
gRNA-3/pATF-3/BFP
gRNA-4/pATF-4/RFP
- dCAS9-VP64
+ dCAS9-VP64
Scale bar : 200 μm
C
Rel. integrated density
con
YFP
BFP
RFP

Figure 5.4

Figure 5.4. Characterization of activity of synthetic pATF promoters.

A. Circuit design of dCas9 based artificial transcription factor-controlled activation of synthetic promoters (pATFs). Specific *gRNAs* are produced by U6 promoter while the expression of the dCas9-VP64 is under the control of the 35S promoter. Reporter genes are under the control of the synthetic promoter (3 repeats of the *gRNA* followed by minimal 35S promoter to the artificial promoter (*gRNA* binding site) upstream of a specific fluorescence reporter. **B.** Fluorescence microscope image showing Agrobacterium mediated transient expression of YFP, BFP and RFP into *Nicotiana benthamiana* leaves with dCas9-VP64 (bottom panels) and without dCas9-VP64 (upper panels) using three different *gRNAs*. Images were captured using the appropriate filter (Materials and Methods) at same exposure. **C.** Relative integrated density of each fluorescence signal (shown in panel B). Integrated density was measured using image J software and normalized to that of the control (con; - dCAS9-VP64). Error bars: S.D. (n=3, independent replicates). Asterisks indicate statistical significance in a student t-test ($P<0.05$).

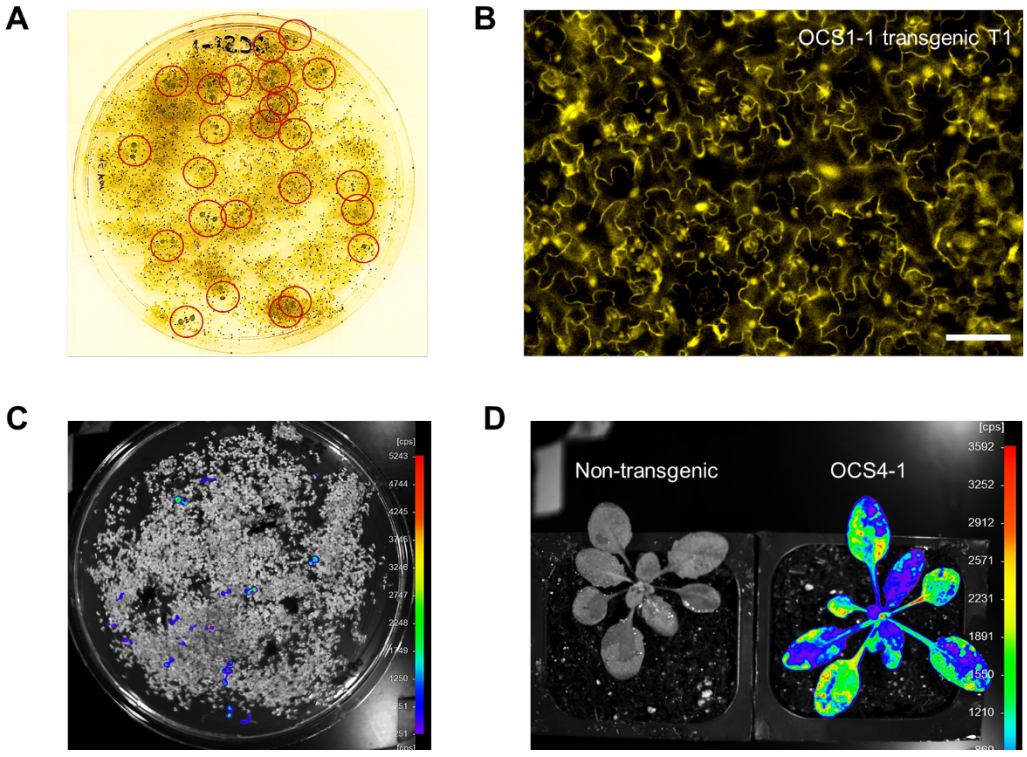

Figure 5.5. Evaluation of OCS reporter gene expression in transgenic Arabidopsis plants.

A. Image showing Kanamycin selection of the transgenic *Arabidopsis* seedlings on MS media. Seedlings highlighted in the red circle have successfully incorporated OCS circuit. Transformation efficiency is within reasonable ranges (~1%) determined by a simple evaluation of the identified seedlings. **B.** Fluorescence microscope image of *Arabidopsis* transgenic T_1 plants containing the constitutive expression of YFP under the OCS control (OCS 1-1). Scale bar: 50 μm **C.** Image showing Kanamycin selection of the transgenic *Arabidopsis* seedlings on MS media using luminescence reporter (OCS4-1) taken using the NightOwl (Methods). **D.** Image of a T_1 *Arabidopsis* plant containing OCS4-1 at the rosette stage after spraying the luciferin (Methods) containing OCS4-1. This image, taken at the rosette stage using NightOwl after luciferin spray, shows that the luciferase expression is active throughout the adult plant. A non-transgenic plant on the left was used as a negative control in the luminescence reporter assay.

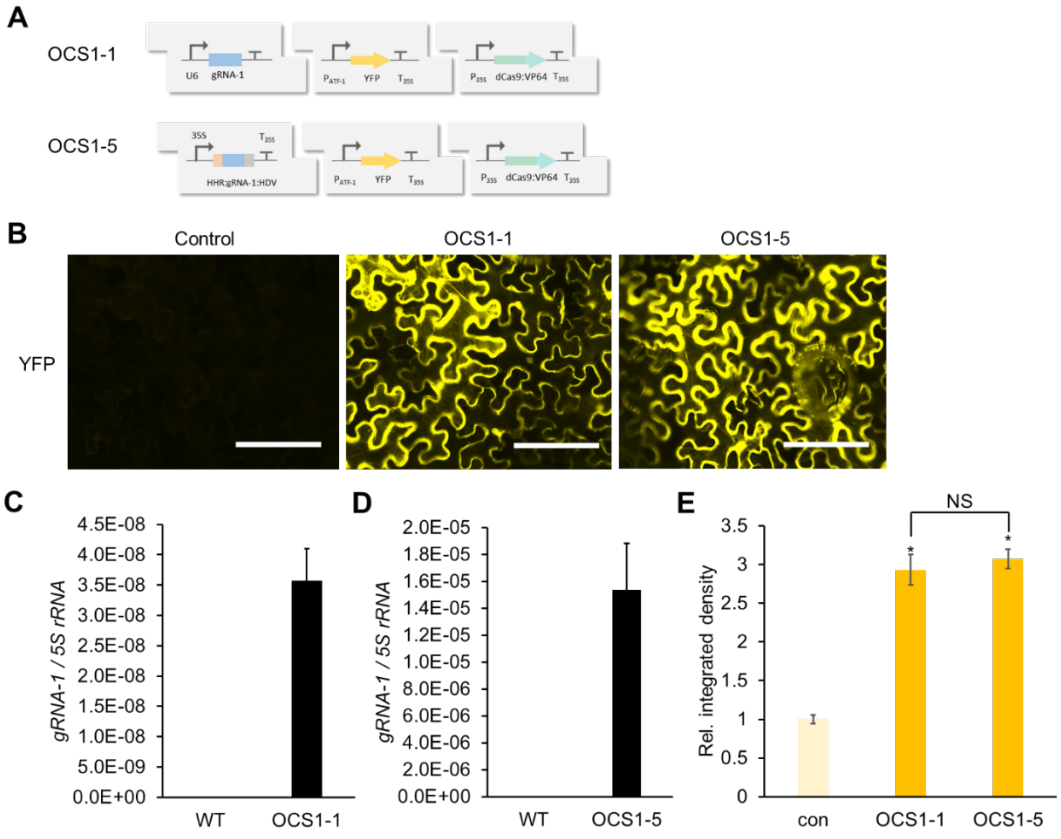

Figure 5.6 Design and characterization of gRNA expression modules under the control of Pol II promoters.

A. OCS1-1 circuit generates RNA using U6 (Pol III) promoter while OCS1-5 circuit generates gRNA using 35S (Pol II) promoter flanked by self-cleaving ribozymes – HammerHead (HHR) and Hepatitis Delta Virus (HDV). **B.** Fluorescence microscope images showing *Agrobacterium* mediated transient expression of OCS constructs with two modalities of gRNA expression (OCS1-1 and OCS1-5). Control images were taken without dCAS9-VP64 expression. Scale bars: 200 µm **C** and **D**. Quantification of the *gRNA-1* expression in OCS constructs (OCS 1-1 (C) and OCS 1-5 (D)) using qPCR relative to *5S rRNA*. Error bars : S.D. (n=3, independent replicates) **E.** Relative integrated density of each fluorescence signal (shown in panel B). Integrated density was measured using image J software and normalized to that of the control (con; - dCas9-VP64). Error bars: S.D. (n=3, independent replicates). Asterisks indicate statistical significance in a student t-test ($P<0.05$). NS: not significant.

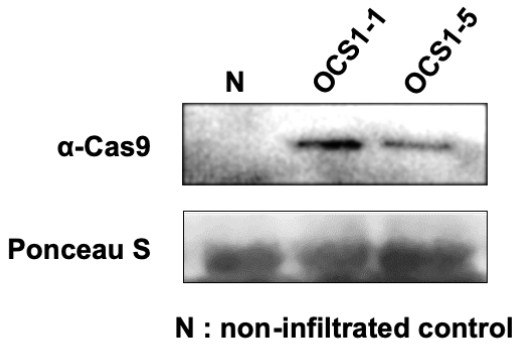

Figure 5.7. The expression of hdCas9 for both OCS 1-1 and OCS 1-5 was confirmed via Western blot analysis (see Materials and Methods).

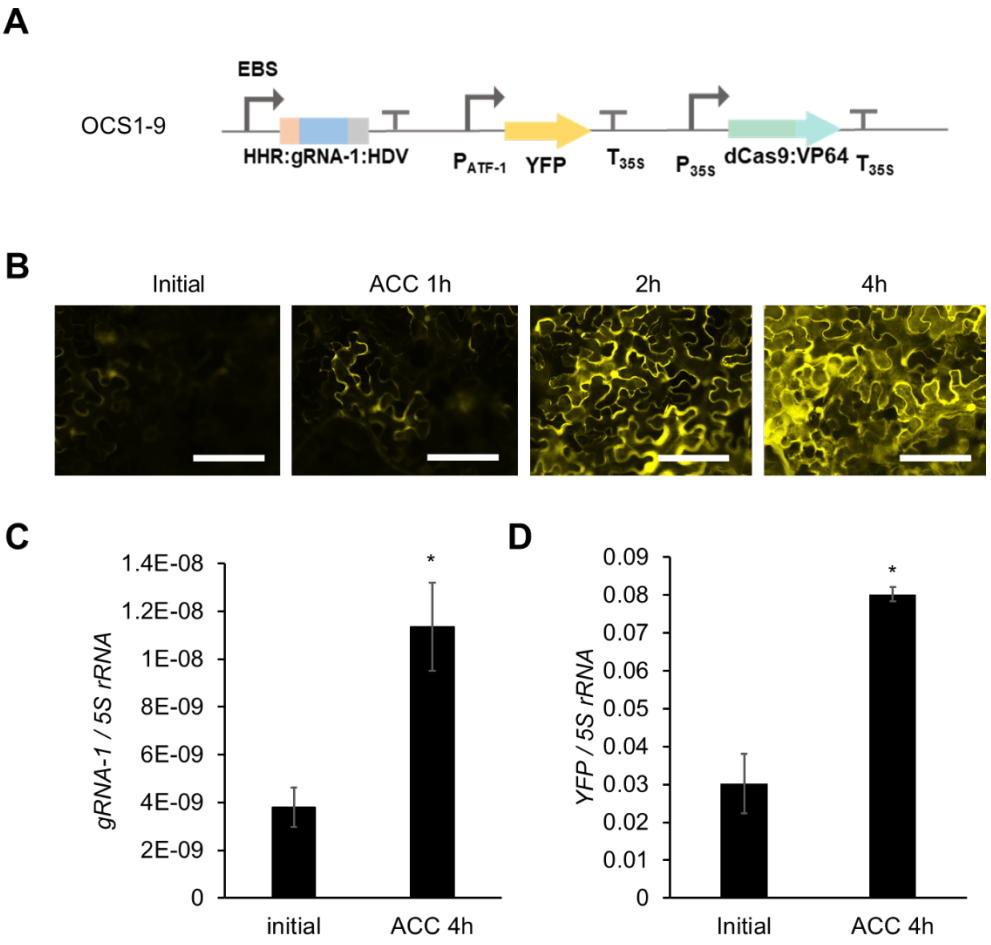

Figure 5.8. Characterization of an ethylene inducible orthogonal control system.

A. OCS1-9 circuit (*gRNA-1* is expressed by ethylene inducible EBS promoter) **B.** Time course fluorescence microscope images showing *Agrobacterium* mediated transient expression of OCS1-9 in *Nicotiana benthamiana* leaves after induction with 10µM ACC. Scale bars: 200 µm **C and D.** qPCR quantification of *gRNA-1* (C) and *YFP* (D) expression before and after induction with ACC, where both show similar levels of induction demonstrating that the relative change in *gRNA-1* expression (ethylene induction) results in the differential activation from the pATF-1 promoter. Error bars: S.D. (n=3, independent replicates), Asterisks indicate statistical significance in a student t-test ($P<0.05$).

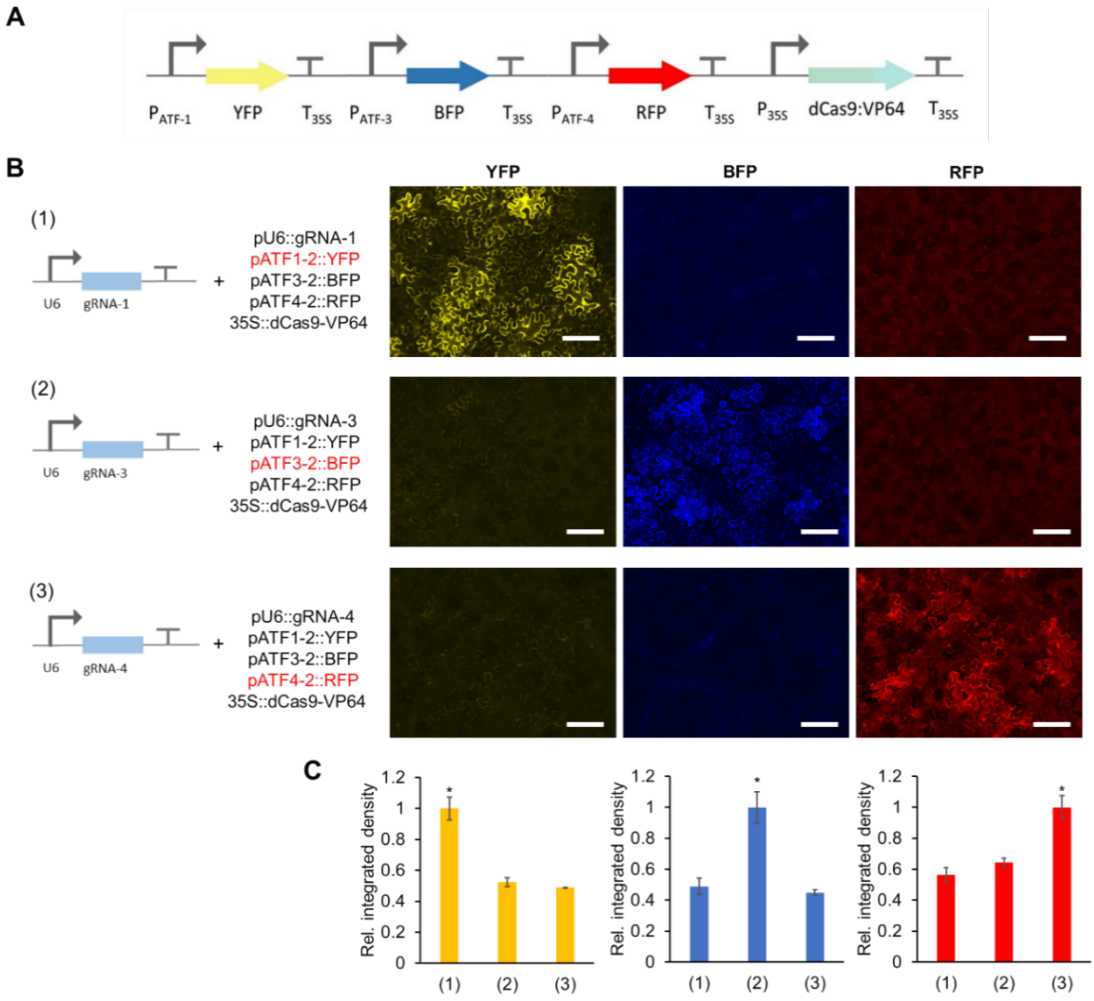

Figure 5.9. Degree of orthogonality of synthetic promoters.

A. OCS circuit containing all three synthetic promoters (pATF-1, pATF-3 and pATF-4) driving three different reporter genes namely YFP, BFP and RFP respectively with a single gRNA expressed one at a time under the control of U6 promoter. **B**. Fluorescence microscope images showing *Agrobacterium* mediated transient expression of OCS constructs in *Nicotiana benthamiana* leaves. Scale bars: 200 µm **C.** As observed from the fluorescence images, only the specific gRNA:pATF pair is active, thus demonstrating that the synthetic promoters are mutually orthogonal Relative integrated density of each fluorescence signal (shown in panel B). Integrated density was measured by image J software and normalized to the highest value. Error bars: S.D. (n=3, independent replicates). Asterisks indicate statistical significance in a student t-test ($P<0.05$).

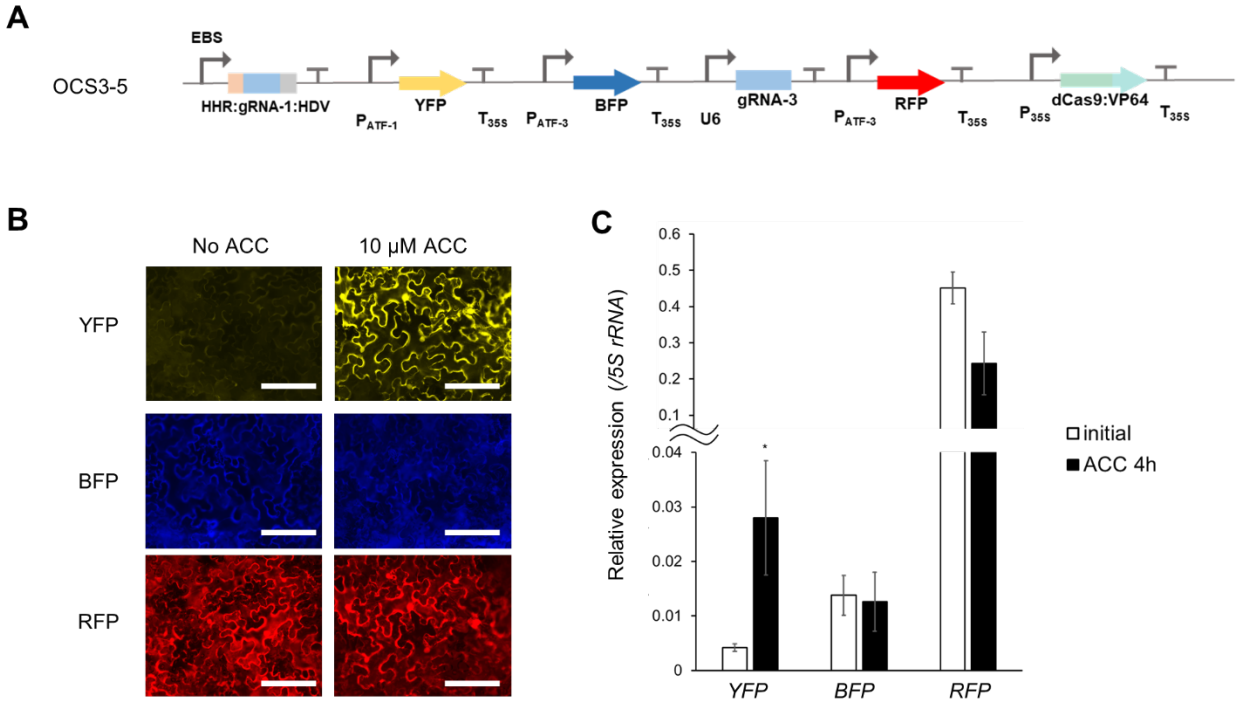

Figure 5.10. Design and characterization of a ratiometric circuit.

A. OSC3-5 contains YFP which is inducible by ACC (pATF-1), while BFP and RFP are constitutively expressed under the control of pATF-3 via the constitutive expression of gRNA-3. **B**. Fluorescence microscope images showing *Agrobacterium* mediated transient expression of the ratiometric OCS construct (OCS3-5) in *Nicotiana benthamiana* leaves with or without 10µM ACC. Scale bars: 200 µm **C.** qPCR quantification of *YFP*, *BFP* and *RFP* shows that YFP is induced after the treatment with ACC while the expression of BFP and RFP remains unchanged before or after ACC induction. Error bars: S.D. (n=4, independent replicates). An asterisk indicates statistical significance in a student t-test

List of genetic elements used**.**

Highlighted regions in red indicate BsaI recognition sites and the corresponding overhangs are highlighted in blue.

Table 5.1: Promoters

Name	Description	Sequence
P1	35S promoter	GGTCTCAAACGCGTCAACATGGTGGAGCACGACACT CTGGTCTACTCCAAAAATGTCAAAGATACAGTCTCAG AAGATCAAAGGGCTATTGAGACTTTTCAACAAAGGA TAATTTCGGGAAACCTCCTCGGATTCCATTGCCCAGC TATCTGTCACTTCATCGAAAGGACAGTAGAAAAGGA AGGTGGCTCCTACAAATGCCATCATTGCGATAAAGG AAAGGCTATCATTCAAGATCTCTCTGCCGACAGTGGT CCCAAAGATGGACCCCCACCCACGAGGAGCATCGTG GAAAAAGAAGAGGTTCCAACCACGTCTACAAAGCAA GTGGATTGATGTGACATCTCCACTGACGTAAGGGATG ACGCACAATCCCACTATCCTTCGCAAGACCCTTCCTC TATATAAGGAAGTTCATTTCATTTGGAGAGGACACGC GTTTATTTACAAGAGCGTACGGTTCAATCCCTGCCTC CCCTGTAAAACTACCCTTTGAAAACCTCTCTTTCTTA ATCTTTTCTTTGTAATTCCAGATCTATGTGAGACC
P2	Mas promoter	GGTCTCAAACGCGGAGATTTTTCAAATCAGTGCGCAA GACGTGACGTAAGTATCCGAGTCAGTTTTTATTTTTC TACTAATTTGGTCGTTTATTTCGGCGTGTAGGACATG GCAACCGGGCCTGAATTTCGCGGGTATTCTGTTTCTA TTCCAACTTTTTCTTGATCCGCAGCCATTAACGACTTT TGAATAGATACGCTGACACGCCAAGCCTCGCTAGTC AAAAGTGTACCAAACAACGCTTTACAGCAAGAACGG AATGCGCGTGACGCTCGCGGTGACGCCATTTCGCCTT TTCAGAAATGGATAAATAGCCTTGCTTCCTATTATAT CTTCCCAAATTACCAATACATTACACTAGCATCTGAA TTTCATAACCAATCTCGATACACCAAATCGAGATCTA TGTGAGACC

Table 5.1 (continued)

P4	EBS promoter	GGTCTCAAACGCATTACGAATTCCCGGGGATCCTCAT GATCAAAGGGGGGATGCACTATTTAAGATCCTCATG ATCAAAGGGGGGATGCACTATTTAAGATCCTCATGAT CAAAGGGGGGATGCACTATTTAAGATCCTCATGATC AAAGGGGGGATGCACTATTTAAGATCCTCATGATCA AAGGGGGGATGCACTATTTAAGATCCTTCGCAAGAC CCTTCCTCTATATAAGGAAGTTCATTTCATTTGGAGA GGACAGGGTATCAAGCTTGGCACTGGCCGTCGTTTTA CAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAA CTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCT GGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTT CCCAACAGTTGCGCAGCCTGAAGCCTAGGGAGGAGT CCACTCTATGTGAGACC
pATF-1	Artificial promoter 1 (3 gRNA repeats)	GGTCTCAAACGCCAAACCACAATTTGCACACCCTGGC ATCTCAAACCACAATTTGCACACCCTGGCATCTCAAA CCACAATTTGCACACCCTGGCATCTCCGCAAGACCCT TCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGA CACGCGTTTATTTACAAGAGCGTACGGTTCAATCCCT GCCTCCCCTGTAAAACTACCCTTTGAAAACCTCTCTT TCTTAATCTTTTCTTTGTAATTCCAGATCTATGTGAGA CC
pATF-3	Artificial promoter 3 (3 gRNA repeats)	GGTCTCAAACGCCTGACTCACAGTCCTATCGAGTGGC ATCTCTGACTCACAGTCCTATCGAGTGGCATCTCTGA CTCACAGTCCTATCGAGTGGCATCTCCGCAAGACCCT TCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGA CACGCGTTTATTTACAAGAGCGTACGGTTCAATCCCT GCCTCCCCTGTAAAACTACCCTTTGAAAACCTCTCTT TCTTAATCTTTTCTTTGTAATTCCAGATCTATGTGAGA CC
pATF-4	Artificial promoter 4 (3 gRNA repeats)	GGTCTCAAACGCCATTGCGCACCATTCCACTAGTGGC ATCTCATTGCGCACCATTCCACTAGTGGCATCTCATT GCGCACCATTCCACTAGTGGCATCTCCGCAAGACCCT TCCTCTATATAAGGAAGTTCATTTCATTTGGAGAGGA CACGCGTTTATTTACAAGAGCGTACGGTTCAATCCCT GCCTCCCCTGTAAAACTACCCTTTGAAAACCTCTCTT TCTTAATCTTTTCTTTGTAATTCCAGATCTATGTGAGA CC

Table 5.2: Terminators

Name	Description	Sequence
T1	35S Terminator	GGTCTCAATCCTAACTCGAGCTCTAGCTAGAGTCGAT CGACAAGCTCGAGTTTCTCCATAATAATGTGTGAGTA GTTCCCAGATAAGGGAATTAGGGTTCCTATAGGGTTT CGCTCATGTGTTGAGCATATAAGAAACCCTTAGTATG TATTTGTATTTGTAAAATACTTCTATCAATAAAATTTC TAATTCCTAAAACCAAAATCCAGTACTAAAATCCAG ATGCTGTGAGACC

Table 5.3 Coding sequences

Name	Description	Sequence
G1	YFP	GGTCTCATATGGTATCCAAAGGAGAAGAATTGTTTAC AGGGGTAGTCCCCATATTGGTCGAGCTCGATGGGGA TGTAAACGGGCACAAGTTCTCAGTGTCTGGCGAAGG CGAAGGGGACGCCACATACGGAAAGTTGACTCTGAA GTTCATCTGCACAACCGGTAAACTACCCGTACCCTGG CCAACACTGGTGACGACCTTTGGGTACGGGCTACAAT GTTTCGCTCGATATCCTGACCACATGAAGCAACACGA TTTCTTCAAAAGCGCCATGCCTGAGGGGTATGTACAG GAGCGTACCATATTTTTCAAAGACGATGGAAACTATA AGACCCGAGCTGAAGTGAAATTTGAAGGAGATACGT TAGTCAATCGAATAGAACTAAAGGGAATTGATTTCA AGGAGGATGGGAATATTCTGGGCCATAAGCTGGAGT ACAATTACAATAGTCACAACGTCTATATAATGGCAG ACAAGCAGAAGAATGGAATAAAGGTTAATTTCAAGA TAAGGCATAACATTGAAGACGGAAGCGTTCAGCTAG CCGATCATTACCAACAGAATACACCGATAGGTGATG GACCCGTCCTGCTCCCCGACAACCATTACTTGTCTTA CCAGAGTGCACTTTCTAAAGATCCAAATGAGAAAAG AGATCATATGGTTTTATTAGAATTTGTTACCGCAGCC GGCATAACTTTGGGAATGGACGAACTGTACAAATGA GGATCCTGAGACC
G2	RFP	GGTCTCATATGAGCGAGCTGATTAAGGAGAACATGC ACATGAAGCTGTACATGGAGGGCACCGTGAACAACC ACCACTTCAAGTGCACATCCGAGGGCGAAGGCAAGC CCTACGAGGGCACCCAGACCATGAGAATCAAGGTGG TCGAGGGCGGCCCTCTCCCCTTCGCCTTCGACATCCT

Table 5.3 (continued)

		GGCTACCAGCTTCATGTACGGCAGCAGAACCTTCATC AACCACACCCAGGGCATCCCCGACTTCTTTAAGCAGT CCTTCCCTGAGGGCTTCACATGGGAGAGAGTCACCAC ATACGAAGACGGGGGCGTGCTGACCGCTACCCAGGA CACCAGCCTCCAGGACGGCTGCCTCATCTACAACGTC AAGATCAGAGGGGTGAACTTCCCATCCAACGGCCCT GTGATGCAGAAGAAAACACTCGGCTGGGAGGCCAAC ACCGAGATGCTGTACCCCGCTGACGGCGGCCTGGAA GGCAGAAGCGACATGGCCCTGAAGCTCGTGGGCGGG GGCCACCTGATCTGCAACTTCAAGACCACATACAGAT CCAAGAAACCCGCTAAGAACCTCAAGATGCCCGGCG TCTACTATGTGGACCACAGACTGGAAAGAATCAAGG AGGCCGACAAAGAAACTTACGTCGAGCAGCATGAGG TTGCTGTGGCCAGATACTGCGACCTCCCTAGCAAACT GGGGCACAAGTGAGGATCCTGAGACC
G10	BFP	GGTCTCATATGAGCGAAGAACTAATCAAGGAAAATA TGCACATGAAACTCTACATGGAGGGTACGGTTGACA ATCATCATTTCAAATGTACCGCAGAGGGCGAAGGTA AACCCTATGAGGGCACCCAGACTATGCGAATCAAGG TTGTGGAGGGCGGGCCATTGCCCTTCGCTTTTGACAT CTTAGCTACTAGTTTCTTATATGGGAGCAAGACCTTT ATAGATCACACACAGGGTATCCCGGACTTCTTTAAAC AGAGCTTTCCAGAAGGGTTCACCTGGGAAAGGGTGA CAACCTATGAAGATGGCGGAGTGCTTACAGCGACAC AAGACACCTCCCTACAGGACGGCACCCTAATATATA ATGTCAAGATTCGTGGCGTAGATTTCCTCAGCAATGG CCCTGTTATGCAGAAGAAGACACTTGGATGGGAGGC TTTCACTGAAACCCTCTACCCGGCGGATGGAGGCTTA GAGGGGAGAAACGATATGGCGTTAAAGCTGGTCGGG GGATCACACTTGATCGCGCATGCAAAAACTACGTAC AGGTCCAAGAAACCAGCAAAGAATCTCAAGATGCCA GGTGTATACTATGTGGATTACCGACTCGAGCGTATTA AGGAAGCAAACGACGAAACGTACGTCGAACAACACG AAGTTGCAGTAGCAAGGTATAGCGACCTTCCCTCCAA GCTAGGACATAAGCTGAATGGGAGCGGATAAGGATC CTGAGACC

Table 5.3 (continued)

G17	Luciferase	GGTCTCATATGGAAGATGCAAAGAATATCAAAAAAG GCCCAGCGCCCTTCTACCCATTAGAAGATGGAACCGC AGGAGAGCAACTTCACAAGGCGATGAAACGATATGC TCTTGTCCCGGGAACCATCGCTTTTACGGACGCACAC ATAGAGGTTAACATTACCTATGCGGAATATTTTGAAA TGTCAGTCAGATTAGCAGAAGCAATGAAACGTTATG GGCTCAACACTAACCATCGTATTGTTGTATGTAGCGA AAACAGCCTGCAGTTTTTTATGCCGGTGCTCGGTGCG CTGTTCATCGGTGTAGCGGTTGCTCCCGCAAACGACA TTTACAACGAGAGAGAACTGCTTAACAGCATGAATA TCAGCCAACCGACCGTCGTGTTTGTCTCAAAAAAGGG ACTACAAAAAATTCTAAATGTCCAAAAGAAGTTACC TATTATCCAGAAAATTATTATTATGGATAGCAAGACG GATTATCAAGGATTCCAATCTATGTACACATTTGTTA CGAGCCACTTACCTCCAGGTTTTAACGAATATGATTT TGTGCCTGAGAGCTTTGATCGAGATAAGACCATCGCG TTAATTATGAATAGTTCCGGCTCTACGGGGCTCCCAA AGGGAGTCGCACTACCACATCGAACTGCGTGCGTTA GATTTTCACATGCCAGAGATCCTATCTTCGGGAATCA GATTATTCCGGACACTGCAATACTGAGTGTGGTTCCG TTTCATCACGGGTTCGGGATGTTCACGACACTCGGTT ACCTCATATGCGGATTTCGTGTGGTGCTGATGTATAG GTTTGAAGAAGAGTTGTTCCTAAGATCCTTGCAGGAT TACAAAATTCAGTCCGCCCTGTTAGTTCCTACCTTATT TTCCTTTTTCGCCAAGTCAACGTTAATTGATAAATAT GACCTATCCAACCTCCACGAAATTGCCAGTGGTGGG GCCCCCTTGTCCAAAGAAGTTGGTGAGGCAGTCGCTA AGAGGTTCCACCTGCCGGGTATCCGTCAAGGCTACG GACTTACCGAAACAACCTCCGCTATTCTTATTACACC TGAAGGCGATGATAAGCCGGGTGCTGTCGGTAAGGT GGTGCCTTTTTTTGAGGCGAAGGTAGTGGACTTAGAT ACTGGCAAGACGCTCGGAGTTAATCAACGAGGCGAG CTCTGTGTCCGTGGTCCCATGATAATGAGCGGATACG TCAACAATCCTGAGGCAACCAACGCATTAATTGATA AAGACGGCTGGTTGCATAGCGGGGATATCGCGTATT GGGATGAGGACGAACACTTCTTTATAGTCGATAGGTT AAAGTCACTTATTAAGTATAAAGGTTATCAAGTCGCT CCCGCCGAACTGGAGAGTATTCTTCTTCAGCATCCGA ATATCTTTGACGCCGGTGTTGCAGGTCTTCCTGACGA CGACGCCGGAGAACTACCTGCAGCGGTCGTGGTGCT

	Table 5.3 (continued)	CGAACATGGAAAGACAATGACCGAGAAGGAGATTGT AGATTACGTAGCTTCACAGGTCACCACCGCTAAAAA ACTAAGAGGTGGTGTTGTCTTCGTAGATGAAGTGCCT AAAGGACTTACCGGTAAACTCGACGCCAGGAAGATA AGAGAAATCCTGATCAAAGCGAAGAAGGGTGGTAAG TCTAAGCTGTAAGGATCCTGAGACC
G15	dCas9:VP64	GGTCTCATATGCCCAAGAAGAAGAGGAAGGTGGACA AGAAGTACTCCATTGGGCTCGCTATCGGCACAAACA GCGTCGGCTGGGCCGTCATTACGGACGAGTACAAGG TGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCG ATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCC TCCTGTTCGACTCCGGGGAAACGGCCGAAGCCACGC GGCTCAAAAGAACAGCACGGCGCAGATATACCCGCA GAAAGAATCGGATCTGCTACCTGCAGGAGATCTTTA GTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCA TAGGCTGGAGGAGTCCTTTTTGGTGGAGGAGGATAA AAAGCACGAGCGCCACCCAATCTTTGGCAATATCGT GGACGAGGTGGCGTACCATGAAAAGTACCCAACCAT ATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGA TAAGGCTGACTTGCGGTTGATCTATCTCGCGCTGGCG CATATGATCAAATTTCGGGGACACTTCCTCATCGAGG GGGACCTGAACCCAGACAACAGCGATGTCGACAAAC TCTTTATCCAACTGGTTCAGACTTACAATCAGCTTTTC GAAGAGAACCCGATCAACGCATCCGGAGTTGACGCC AAAGCAATCCTGAGCGCTAGGCTGTCCAAATCCCGG CGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAG AAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGT CACTCGGGCTGACCCCCAACTTTAAATCTAACTTCGA CCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGA CACCTACGATGATGATCTCGACAATCTGCTGGCCCAG ATCGGCGACCAGTACGCAGACCTTTTTTTGGCGGCAA AGAACCTGTCAGACGCCATTCTGCTGAGTGATATTCT GCGAGTGAACACGGAGATCACCAAAGCTCCGCTGAG CGCTAGTATGATCAAGCGCTATGATGAGCACCACCA AGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAA CTGCCTGAGAAGTACAAGGAAATTTTCTTCGATCAGT CTAAAAATGGCTACGCCGGATACATTGACGGCGGAG CAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCAT CTTGGAAAAAATGGACGGCACCGAGGAGCTGCTGGT

Table 5.3 (continued)

AAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCG
CACTTTCGACAATGGAAGCATCCCCCACCAGATTCAC
CTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAG
GATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGA
TTGAGAAAATCCTCACATTTCGGATACCCTACTATGT
AGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTG
GATGACTCGCAAATCAGAAGAGACTATCACTCCCTG
GAACTTCGAGGAAGTCGTGGATAAGGGGGCCTCTGC
CCAGTCCTTCATCGAAAGGATGACTAACTTTGATAAA
AATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTC
TGCTGTACGAGTACTTCACAGTTTATAACGAGCTCAC
CAAGGTCAAATACGTCACAGAAGGGATGAGAAAGCC
AGCATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGT
GGACCTCCTCTTCAAGACGAACCGGAAAGTTACCGT
GAAACAGCTCAAAGAAGATTATTTCAAAAAGATTGA
ATGTTTCGACTCTGTTGAAATCAGCGGAGTGGAGGAT
CGCTTCAACGCATCCCTGGGAACGTATCACGATCTCC
TGAAAATCATTAAAGACAAGGACTTCCTGGACAATG
AGGAGAACGAGGACATTCTTGAGGACATTGTCCTCA
CCCTTACGTTGTTTGAAGATAGGGAGATGATTGAAGA
ACGCTTGAAAACTTACGCTCATCTCTTCGACGACAAA
GTCATGAAACAGCTCAAGAGGCGCCGATATACAGGA
TGGGGGCGGCTGTCAAGAAAACTGATCAATGGGATC
CGAGACAAGCAGAGTGGAAAGACAATCCTGGATTTT
CTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGC
AGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGA
CATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAG
TCTCCACGAGCACATCGCTAATCTTGCAGGTAGCCCA
GCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTC
GTGGATGAACTCGTCAAAGTAATGGGAAGGCATAAG
CCCGAGAATATCGTTATCGAGATGGCCCGAGAGAAC
CAAACTACCCAGAAGGGACAGAAGAACAGTAGGGA
AAGGATGAAGAGGATTGAAGAGGGTATAAAAGAACT
GGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAA
CACCCAGCTTCAGAATGAGAAGCTCTACCTGTACTAC
CTGCAGAACGGCAGGGACATGTACGTGGATCAGGAA
CTGGACATCAATCGGCTCTCCGACTACGACGTGGATG
CCATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATT
GATAATAAAGTGTTGACAAGATCCGATAAAAATAGA
GGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTC

Table 5.3 (continued)

AAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAAC
GCCAAACTGATCACACAACGGAAGTTCGATAATCTG
ACTAAGGCTGAACGAGGTGGCCTGTCTGAGTTGGAT
AAAGCCGGCTTCATCAAAAGGCAGCTTGTTGAGACA
CGCCAGATCACCAAGCACGTGGCCCAAATTCTCGATT
CACGCATGAACACCAAGTACGATGAAAATGACAAAC
TGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAA
GCTGGTTTCAGATTTCAGAAAGGACTTTCAGTTTTAT
AAGGTGAGAGAGATCAACAATTACCACCATGCGCAT
GATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTA
TCAAAAAATATCCCAAGCTTGAATCTGAATTTGTTTA
CGGAGACTATAAAGTGTACGATGTTAGGAAAATGAT
CGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGC
TAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCA
AGACCGAGATTACACTGGCCAATGGAGAGATTCGGA
AGCGACCACTTATCGAAACAAACGGAGAAACAGGAG
AAATCGTGTGGGACAAGGGTAGGGATTTCGCGACAG
TCCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGT
TAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAA
GGAAAGTATCCTCCCGAAAAGGAACAGCGACAAGCT
GATCGCACGCAAAAAAGATTGGGACCCCAAGAAATA
CGGCGGATTCGATTCTCCTACAGTCGCTTACAGTGTA
CTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAA
AAACTCAAAAGCGTCAAGGAACTGCTGGGCATCACA
ATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATC
GACTTTCTCGAGGCGAAAGGATATAAAGAGGTCAAA
AAAGACCTCATCATTAAGCTTCCCAAGTACTCTCTCT
TTGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTA
GTGCGGGCGAGCTGCAGAAAGGTAACGAGCTGGCAC
TGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAG
CCACTATGAAAAGCTCAAAGGATCTCCCGAAGATAA
TGAGCAGAAGCAGCTGTTCGTGGAACAACACAAACA
CTACCTTGATGAGATCATCGAGCAAATAAGCGAATTC
TCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATA
AGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCC
CATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTT
ACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAAGT
ACTTCGACACCACCATAGACAGAAAGCGGTACACCT
CTACAAAGGAGGTCCTGGACGCCACACTGATTCATC
AGTCAATTACGGGGCTCTATGAAACAAGAATCGACC

Table 5.3 (continued)		TCTCTCAGCTCGGTGGAGACAGCAGGGCTGATTCGG ACCCAAAAAAGAAGCGTAAGGTCGATCCCAAGAAGA AGAGAAAGGTAGATCCTAAGAAGAAGAGAAAGGTA GACGCATTGGATGACTTCGACTTAGACATGCTTGGGT CAGATGCTTTGGACGACTTTGACCTCGACATGTTGGG TTCTGACGCGCTAGATGACTTCGACTTGGACATGTTG GGCTCCGACGCTTTGGACGACTTTGATCTGGACATGT TATGAGGATCCTGAGACC

Table 5.4: gRNA expression cassettes

Name	Description	Sequence
U6-gRNA-1	U6 promoter driving gRNA-1 expression	GGTCTCAAACGTTCGACGTAAAGCCTGTAGAAGAGG TTTCTAGCGAACGACACGAGTTTGAGCCTCATGAAGC TTCGTTGAACAACGGAAACTCGACTTGCCTTCCGCAC AATACATCATTTCTTCTTAGCTTTTTTTCTTCTTCTTCG TTCATACAGTTTTTTTTTGTTTATCAGCTTACATTTTCT TGAACCGTAGCTTTCGTTTTCTTCTTTTTAACTTTCCA TTCGGAGTTTTTGTATCTTGTTTCATAGTTTGTCCCAG GATTAGAATGATTAGGCATCGAACCTTCAAGAATTTG ATTGAATAAAACATCTTCATTCTTAAGATATGAAGAT AATCTTCAAAAGGCCCCTGGGAATCTGAAAGAAGAG AAGCAGGCCCATTTATATGGGAAAGAACAATAGTAT TTCTTATATAGGCCCATTTAAGTTGAAAACAATCTTC AAAAGTCCCACATCGCTTAGATAAGAAAACGAAGCT GAGTTTATATACAGCTAGAGTCGAAGTAGTGATTAA ACCACAATTTGCACACCCGTTTTAGAGCTAGAAATAG CAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTTTTGCAAAATTTT CCAGATCGATTTCTTCTTCCTCTGTTCTTCGGCGTTCA ATTTCTGGGTTTTTCTCTTCGTTTTCTGTAACTGAAAC CTAAAATTTGACCTAAAAAAAATCTCAAATAATATG ATTCAGTGGTTTTGTACTTTTCAGTTAGTTGAGTTTTG CAGTTCCGATGAGATAAACCAATAACTTTGCTTAGAT CTAATTCATTCCGTTACACCTCTGATGGAGATGGAAG GTTCTTAATAATGATGCCATTTTTTGGGTAATAATTTT GAATTAGAATCAAGGGTATAAGATTCATAATTAACA

Table 5.4 (continued)

		TCACTTAAGCAAAGTTCGTAATATACGACCACAGGAT ATAATTTTTGGTACGCTGTGAGACC
U6-gRNA-3	U6 promoter driving gRNA-3 expression	GGTCTCAAACGTTCGACGTAAAGCCTGTAGAAGAGG TTTCTAGCGAACGACACGAGTTTGAGCCTCATGAAGC TTCGTTGAACAACGGAAACTCGACTTGCCTTCCGCAC AATACATCATTTCTTCTTAGCTTTTTTTCTTCTTCTTCG TTCATACAGTTTTTTTTTGTTTATCAGCTTACATTTTCT TGAACCGTAGCTTTCGTTTTCTTCTTTTTAACTTTCCA TTCGGAGTTTTTGTATCTTGTTTCATAGTTTGTCCCAG GATTAGAATGATTAGGCATCGAACCTTCAAGAATTTG ATTGAATAAAACATCTTCATTCTTAAGATATGAAGAT AATCTTCAAAAGGCCCCTGGGAATCTGAAAGAAGAG AAGCAGGCCCATTTATATGGGAAAGAACAATAGTAT TTCTTATATAGGCCCATTTAAGTTGAAAACAATCTTC AAAAGTCCCACATCGCTTAGATAAGAAAACGAAGCT GAGTTTATATACAGCTAGAGTCGAAGTAGTGATTTGA CTCACAGTCCTATCGAGGTTTTAGAGCTAGAAATAGC AAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTTTTGCAAAATTTT CCAGATCGATTTCTTCTTCCTCTGTTCTTCGGCGTTCA ATTTCTGGGTTTTTCTCTTCGTTTTCTGTAACTGAAAC CTAAAATTTGACCTAAAAAAAATCTCAAATAATATG ATTCAGTGGTTTTGTACTTTTCAGTTAGTTGAGTTTTG CAGTTCCGATGAGATAAACCAATAACTTTGCTTAGAT CTAATTCATTCCGTTACACCTCTGATGGAGATGGAAG GTTCTTAATAATGATGCCATTTTTTGGGTAATAATTTT GAATTAGAATCAAGGGTATAAGATTCATAATTAACA TCACTTAAGCAAAGTTCGTAATATACGACCACAGGAT ATAATTTTTGGTACGCTGTGAGACC
U6-gRNA-4	U6 promoter driving gRNA-4 expression	GGTCTCAAACGTTCGACGTAAAGCCTGTAGAAGAGG TTTCTAGCGAACGACACGAGTTTGAGCCTCATGAAGC TTCGTTGAACAACGGAAACTCGACTTGCCTTCCGCAC AATACATCATTTCTTCTTAGCTTTTTTTCTTCTTCTTCG TTCATACAGTTTTTTTTTGTTTATCAGCTTACATTTTCT TGAACCGTAGCTTTCGTTTTCTTCTTTTTAACTTTCCA TTCGGAGTTTTTGTATCTTGTTTCATAGTTTGTCCCAG GATTAGAATGATTAGGCATCGAACCTTCAAGAATTTG ATTGAATAAAACATCTTCATTCTTAAGATATGAAGAT AATCTTCAAAAGGCCCCTGGGAATCTGAAAGAAGAG AAGCAGGCCCATTTATATGGGAAAGAACAATAGTAT

Table 5.4 (continued)

		TTCTTATATAGGCCCATTTAAGTTGAAAACAATCTTC AAAAGTCCCACATCGCTTAGATAAGAAAACGAAGCT GAGTTTATATACAGCTAGAGTCGAAGTAGTGATTATT GCGCACCATTCCACTAGGTTTTAGAGCTAGAAATAGC AAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAA AAGTGGCACCGAGTCGGTGCTTTTTTTGCAAAATTTT CCAGATCGATTTCTTCTTCCTCTGTTCTTCGGCGTTCA ATTTCTGGGTTTTTCTCTTCGTTTTCTGTAACTGAAAC CTAAAATTTGACCTAAAAAAAATCTCAAATAATATG ATTCAGTGGTTTTGTACTTTTCAGTTAGTTGAGTTTTG CAGTTCCGATGAGATAAACCAATAACTTTGCTTAGAT CTAATTCATTCCGTTACACCTCTGATGGAGATGGAAG GTTCTTAATAATGATGCCATTTTTTGGGTAATAATTTT GAATTAGAATCAAGGGTATAAGATTCATAATTAACA TCACTTAAGCAAAGTTCGTAATATACGACCACAGGAT ATAATTTTTGGTACGCTGTGAGACC
HHR-gRNA-1	gRNA-1 expression flanked by 5’ HHR and 3’ HDV	GGTCTCATATGCGACTACTGATGAGTCCGTGAGGACG AAACGAGTAAGCTCGTCAAACCACAATTTGCACACC CGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC TAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG GTGCTTTTGGCCGGCATGGTCCCAGCCTCCTCGCTGG CGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGG GACGGATCCTGAGACC
HHR-gRNA-3	gRNA-3 expression flanked by 5’ HHR and 3’ HDV	GGTCTCATATGGAGTCACTGATGAGTCCGTGAGGAC GAAACGAGTAAGCTCGTCTGACTCACAGTCCTATCGA GGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC TAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG GTGCTTTTGGCCGGCATGGTCCCAGCCTCCTCGCTGG CGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGG GACGGATCCTGAGACC
HHR-gRNA-4	gRNA-4 expression flanked by 5’ HHR and 3’ HDV	GGTCTCATATGCGCAATCTGATGAGTCCGTGAGGACG AAACGAGTAAGCTCGTCATTGCGCACCATTCCACTAG GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCT AGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCG GTGCTTTTGGCCGGCATGGTCCCAGCCTCCTCGCTGG CGCCGGCTGGGCAACATGCTTCGGCATGGCGAATGG GACGGATCCTGAGACC

Table 5.5: Plant expression vector backbone elements

Name	Description	Sequence
Pnos	Nos promoter driving the plant resistance cassette	AGCGGAGAATTAAGGGAGTCACGTTATGACCCCCG CCGATGACGCGGGACAAGCCGTTTTACGTTTGGAA CTGACAGAACCGCAACGTTGAAGGAGCCACTCAGC CGCGGGTTTCTGGAGTTTAATGAGCTAAGCACATAC GTCAGAAACCATTATTGCGCGTTCAAAAGTCGCCTA AGGTCACTATCAGCTAGCAAATATTTCTTGTCAAAA ATGCTCCACTGACGTTCCATAAATTCCCCTCGGTAT CCAATTAGAGTCTCATATTCACTCTCAATCCAAATA ATCTGCACCGGATCT
KanR (Plant)	Kanamycin coding sequence (plant)	ATGATTGAACAAGATGGATTGCACGCAGGTTCTCC GGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACT GGGCACAACAGACAATCGGCTGCTCTGATGCCGCC GTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTT TTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAA CTCCAGGACGAGGCAGCGCGGCTATCGTGGCTGGC CACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGT TGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGG GCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACC TTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATG CAATGCGGCGGCTGCATACGCTTGATCCGGCTACCT GCCCATTCGACCACCAAGCGAAACATCGCATCGAG CGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGA TCAGGATGATCTGGACGAAGAGCATCAGGGGCTCG CGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGCGT ATGCCCGACGGCGAGGATCTCGTCGTGACTCATGG CGATGCCTGCTTGCCGAATATCATGGTGGAAAATG GCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGG GTGTGGCGGACCGCTATCAGGACATAGCGTTGGCT ACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATG GGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGC TCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTT GACGAGTTCTTC
Tnos	Nos terminator	CTAGAGTCAAGCAGATCGTTCAAACATTTGGCAAT AAAGTTTCTTAAGATTGAATCCTGTTGCCGGTCTTG CGATGATTATCATATAATTTCTGTTGAATTACGTTA AGCATGTAATAATTAACATGTAATGCATGACGTTAT TTATGAGATGGGTTTTTATGATTAGAGTCCCGCAAT TATACATTTAATACGCGATAGAAAACAAAATATAG

Table 5.5 (continued)

		CGCGCAAACTAGGATAAATTATCGCGCGCGGTGTC ATCTATGTTACTAGATCGA
pVS1 replicon	OriV; RepA ; StaA for propagation in Agrobacterium	CGTGCGGCTGCATGAAATCCTGGCCGGTTTGTCTGA TGCCAAGCTGGCGGCCTGGCCGGCCAGCTTGGCCG CTGAAGAAACCGAGCGCCGCCGTCTAAAAAGGTGA TGTGTATTTGAGTAAAACAGCTTGCGTCATGCGGTC GCTGCGTATATGATGCGATGAGTAAATAAACAAAT ACGCAAGGGGAACGCATGAAGGTTATCGCTGTACT TAACCAGAAAGGCGGGTCAGGCAAGACGACCATCG CAACCCATCTAGCCCGCGCCCTGCAACTCGCCGGG GCCGATGTTCTGTTAGTCGATTCCGATCCCCAGGGC AGTGCCCGCGATTGGGCGGCCGTGCGGGAAGATCA ACCGCTAACCGTTGTCGGCATCGACCGCCCGACGAT TGACCGCGACGTGAAGGCCATCGGCCGGCGCGACT TCGTAGTGATCGACGGAGCGCCCCAGGCGGCGGAC TTGGCTGTGTCCGCGATCAAGGCAGCCGACTTCGTG CTGATTCCGGTGCAGCCAAGCCCTTACGACATATGG GCCACCGCCGACCTGGTGGAGCTGGTTAAGCAGCG CATTGAGGTCACGGATGGAAGGCTACAAGCGGCCT TTGTCGTGTCGCGGGCGATCAAAGGCACGCGCATC GGCGGTGAGGTTGCCGAGGCGCTGGCCGGGTACGA GCTGCCCATTCTTGAGTCCCGTATCACGCAGCGCGT GAGCTACCCAGGCACTGCCGCCGCCGGCACAACCG TTCTTGAATCAGAACCCGAGGGCGACGCTGCCCGC GAGGTCCAGGCGCTGGCCGCTGAAATTAAATCAAA ACTCATTTGAGTTAATGAGGTAAAGAGAAAATGAG CAAAAGCACAAACACGCTAAGTGCCGGCCGTCCGA GCGCACGCAGCAGCAAGGCTGCAACGTTGGCCAGC CTGGCAGACACGCCAGCCATGAAGCGGGTCAACTT TCAGTTGCCGGCGGAGGATCACACCAAGCTGAAGA TGTACGCGGTACGCCAAGGCAAGACCATTACCGAG CTGCTATCTGAATACATCGCGCAGCTACCAGAGTAA ATGAGCAAATGAATAAATGAGTAGATGAATTTTAG CGGCTAAAGGAGGCGGCATGGAAAATCAAGAACA ACCAGGCACCGACGCCGTGGAATGCCCCATGTGTG GAGGAACGGGCGGTTGGCCAGGCGTAAGCGGCTGG GTTGTCTGCCGGCCCTGCAATGGCACTGGAACCCCC AAGCCCGAGGAATCGGCGTGACGGTCGCAAACCAT CCGGCCCGGTACAAATCGGCGCGGCGCTGGGTGAT

Table 5.5 (continued)

GACCTGGTGGAGAAGTTGAAGGCCGCGCAGGCCGC
CCAGCGGCAACGCATCGAGGCAGAAGCACGCCCCG
GTGAATCGTGGCAAGCGGCCGCTGATCGAATCCGC
AAAGAATCCCGGCAACCGCCGGCAGCCGGTGCGCC
GTCGATTAGGAAGCCGCCCAAGGGCGACGAGCAAC
CAGATTTTTTCGTTCCGATGCTCTATGACGTGGGCA
CCCGCGATAGTCGCAGCATCATGGACGTGGCCGTTT
TCCGTCTGTCGAAGCGTGACCGACGAGCTGGCGAG
GTGATCCGCTACGAGCTTCCAGACGGGCACGTAGA
GGTTTCCGCAGGGCCGGCCGGCATGGCCAGTGTGT
GGGATTACGACCTGGTACTGATGGCGGTTTCCCATC
TAACCGAATCCATGAACCGATACCGGGAAGGGAAG
GGAGACAAGCCCGGCCGCGTGTTCCGTCCACACGT
TGCGGACGTACTCAAGTTCTGCCGGCGAGCCGATG
GCGGAAAGCAGAAAGACGACCTGGTAGAAACCTGC
ATTCGGTTAAACACCACGCACGTTGCCATGCAGCGT
ACGAAGAAGGCCAAGAACGGCCGCCTGGTGACGGT
ATCCGAGGGTGAAGCCTTGATTAGCCGCTACAAGA
TCGTAAAGAGCGAAACCGGGCGGCCGGAGTACATC
GAGATCGAGCTAGCTGATTGGATGTACCGCGAGAT
CACAGAAGGCAAGAACCCGGACGTGCTGACGGTTC
ACCCCGATTACTTTTTGATCGATCCCGGCATCGGCC
GTTTTCTCTACCGCCTGGCACGCCGCGCCGCAGGCA
AGGCAGAAGCCAGATGGTTGTTCAAGACGATCTAC
GAACGCAGTGGCAGCGCCGGAGAGTTCAAGAAGTT
CTGTTTCACCGTGCGCAAGCTGATCGGGTCAAATGA
CCTGCCGGAGTACGATTTGAAGGAGGAGGCGGGGC
AGGCTGGCCCGATCCTAGTCATGCGCTACCGCAACC
TGATCGAGGGCGAAGCATCCGCCGGTTCCTAATGT
ACGGAGCAGATGCTAGGGCAAATTGCCCTAGCAGG
GGAAAAAGGTCGAAAAAGCTTCTTTCCTGTGGATA
GCACGTACATTGGGAACCCAAAGCCGTACATTGGG
AACCGGAACCCGTACATTGGGAACCCAAAGCCGTA
CATTGGGAACCGGTCACACATGTAAGTGACTGATA
TAAAAGAGAAAAAAGGCGATTTTTCCGCCTAAAAC
TCTTTAAAACTTATTAAAACTCTTAAAACCCGCCTG
GCCTGTGCATAACTGTCTGGCCAGCGCACAGCCGA
AGAGCTGCAAAAAGCGCCTACCCTTCGGTCGCTGC
GCTCCCTACGCCCCGCCGCTTCGCGTCGGCCTATCG
CGGCCGCTGGCCGCTCAAAAATGGCTGGCCTACGG

Table 5.5 (continued)

		CCAGGCAATCTACCAGGGCGCGGACAAGCCGCGCC GTCGCCACTCGACCGCCGGCGCCCACATCAAGGCA CC
pMB1 origin	Origin of replication for propagation in E.coli	AAAGGATCTTCCTGAGATCCTTTTTTTCTGCGCGTA ATCTGCTGCTTGCAAACAAAAAAACCACCGCTACC AGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAAC TCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCA GATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTT AGGCCACCACTTCAAGAACTCTGTAGCACCGCCTAC ATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGC TGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGA CTCAAGACGATAGTTACCGGATAAGGCGCAGCGGT CGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGC TTGGAGCGAACGACCTACACCGAACTGAGATACCT ACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCG AAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGG CAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTC CAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCG GGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGT GATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAAC GCCAGCAACGCG
KanR (bacterial)	Kanamycin expression cassette for bacterial selection	GCCAATTCGTGCGCGGAACCCCTATTTGTTTATTTTT CTAAATACATTCAAATATGTATCCGCTCATGAGACA ATAACCCTGATAAATGCTTCAATAATATTGAAAAA GGAAGAGTATGGCTAAAATGAGAATATCACCGGAA TTGAAAAAACTGATCGAAAAATACCGCTGCGTAAA AGATACGGAAGGAATGTCTCCTGCTAAGGTATATA AGCTGGTGGGAGAAAATGAAAACCTATATTTAAAA ATGACGGACAGCCGGTATAAAGGGACCACCTATGA TGTGGAACGGGAAAAGGACATGATGCTATGGCTGG AAGGAAAGCTGCCTGTTCCAAAGGTCCTGCACTTTG AACGGCATGATGGCTGGAGCAATCTGCTCATGAGT GAGGCCGATGGCGTCCTTTGCTCGGAAGAGTATGA AGATGAACAAAGCCCTGAAAAGATTATCGAGCTGT ATGCGGAGTGCATCAGGCTCTTTCACTCCATCGACA TATCGGATTGTCCCTATACGAATAGCTTAGACAGCC GCTTAGCCGAATTGGATTACTTACTGAATAACGATC TGGCCGATGTGGATTGCGAAAACTGGGAAGAGGAC ACTCCATTTAAAGATCCGCGCGAGCTGTATGATTTT TTAAAGACGGAAAAGCCCGAAGAGGAACTTGTCTT

	Table 5.5 (continued)	TTCCCACGGCGACCTGGGAGACAGCAACATCTTTGT GAAAGATGGCAAAGTAAGTGGCTTTATTGATCTTG GGAGAAGCGGCAGGGCGGACAAGTGGTATGACATT GCCTTCTGCGTCCGGTCGATCAGGGAGGATATCGG GGAAGAACAGTATGTCGAGCTATTTTTTGACTTACT GGGGATCAAGCCTGATTGGGAGAAAATAAAATATT ATATTTTACTGGATGAATTGTTTTAGCTGTCAGACC AAGTTTACTCATATATACTTTAGATTGATTTAAAAC TTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCC TTTTTGATAATC
LB	Left border repeat	CTGATGGGCTGCCTGTATCGAGTGGTGATTTTGTGC CGAGCTGCCGGTCGGGGAGCTGTTGGCTGGCTGGT GGCAGGATATATTGTGGTGTAAACAAATTGACGCTT AGACAACTTAATAACACATTGCGGACGTTTTTAATG TACTG
RB	Right border repeat	TGGTTGGCACATACAAATGGACGAACGGATAAACC TTTTCACGCCCTTTTAAATATCCGATTATTCTAATAA ACGCTCTTTTCTCTTAGGTTTACCCGCCAATATATCC TGTCAAACACTGATAGTTT
GFP DO	GFP expression cassette drop-out flanked by connector sequences	ACTGATGAGACGTGGTAGAGCCACAAACAGCCGGT ACAAGCAACGATCTCCAGGACCATCTGAATCATGC GCGGATGACACGAACTCACGACGGCGATCACAGAC ATTAACCCACAGTACAGACACTGCGACAACGTGGC AATTCGTCGCAATACAACGTGAGACCGAAAGTGAA ACGTGATTTCATGCGTCATTTTGAACATTTTGTAAA TCTTATTTAATAATGTGTGCGGCAATTCACATTTAA TTTATGAATGTTTTCTTAACATCGCGGCAACTCAAG AAACGGCAGGTTCGGATCTTAGCTACTAGAGAAAG AGGAGAAATACTAGATGCGTAAAGGCGAAGAGCTG TTCACTGGTGTCGTCCCTATTCTGGTGGAACTGGAT GGTGATGTCAACGGTCATAAGTTTTCCGTGCGTGGC GAGGGTGAAGGTGACGCAACTAATGGTAAACTGAC GCTGAAGTTCATCTGTACTACTGGTAAACTGCCGGT TCCTTGGCCGACTCTGGTAACGACGCTGACTTATGG TGTTCAGTGCTTTGCTCGTTATCCGGACCATATGAA GCAGCATGACTTCTTCAAGTCCGCCATGCCGGAAG GCTATGTGCAGGAACGCACGATTTCCTTTAAGGATG ACGGCACGTACAAAACGCGTGCGGAAGTGAAATTT GAAGGCGATACCCTGGTAAACCGCATTGAGCTGAA AGGCATTGACTTTAAAGAGGACGGCAATATCCTGG

	Table 5.5 (continued)	
		GCCATAAGCTGGAATACAATTTTAACAGCCACAAT GTTTACATCACCGCCGATAAACAAAAAAATGGCAT TAAAGCGAATTTTAAAATTCGCCATAACGTTGAGG ATGGCAGCGTGCAGCTGGCTGATCACTACCAGCAA AACACTCCAATCGGTGATGGTCCTGTTCTGCTGCCA GACAATCACTATCTGAGCACGCAAAGCGTTCTGTCT AAAGATCCGAACGAGAAACGCGATCATATGGTTCT GCTGGAGTTCGTAACCGCAGCGGGCATCACGCATG GTATGGATGAACTGTACAAATGACCAGGCATCAAA TAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTT CGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCTACT AGAGTCACACTGGCTCACCTTCGGGTGGGCCTTTCT GCGTTTATAGGTCTCAGCTGGAAATCTGCTCGTCAG TGGTGCTCACACTGACGAATCATGTACAGATCATAC CGATGACTGCCTGGCGACTCACAACTAAGCAAGAC AGCCGGAACCAGCGCCGGCGAACACCACTGCATAT ATGGCATATCACAACAGTCCACGTCTCAAGCAGTTA CAGAGATGTTACGAACCG

Table 5.6: List of OCS constructs used

Name	Description	Transcriptional Units
OCS 1-1 (13,694bp)	U6 driven constitutive expression of gRNA-1 leading to constitutive expression of YFP under the control of pATF-1 promoter	TU1 : U6 :gRNA-1 (1013 bp) TU2 : pATF-1: YFP: T35S (1231 bp) TU3E : P35S : dCas9:VP64: T35S (5162 bp)
OCS 1-5 (13,685bp)	35S driven expression of gRNA-1 (flanked by ribozymes) leading to constitutive expression of YFP under the control of pATF-1 promoter	TU1 : P35S: HHR-gRNA-1-HDV:T35S (1001 bp) TU2 : pATF-1 : YFP: T35S (1231 bp) TU3E : P35S : dCas9:VP64: T35S (5162 bp)

Table 5.6 (continued)

OCS 1-9 (13,592bp)	Ethylene inducible expression of gRNA-1 under the control of EBS promoter. YFP under the control of pATF-1 promoter	TU1 : EBS: HHR-gRNA-1-HDV:T35S (917 bp) TU2 : pATF-1 : YFP: T35S (1231 bp) TU3E : P35S : dCas9:VP64: T35S (5162 bp)
OCS 4-1 (14,627bp)	U6 driven constitutive expression of gRNA-1 leading to constitutive expression of Luc under the control of pATF-1 promoter	TU1 : U6 :gRNA-1 (1013 bp) TU2 : pATF-1 : Luc: T35S (2150 bp) TU3E : P35S : dCas9:VP64: T35S (5162 bp)
OCS 3-5 (16,989bp)	Ratiometric circuit where YFP under the control of pATF-1 inducible by ethylene, while RFP and BFP under the control of Patf-3 are constitutively expressed	TU1 : EBS: HHR-gRNA-1-HDV:T35S (917 bp) TU2 : pATF-1 : YFP: T35S (1231 bp) TU3: pATF-3 : BFP : T35S (1228 bp) TU4: U6 :gRNA3 (1013 bp) TU5: pATF-3 : RFP : T35S (1214 bp) TU6E : P35S : dCas9:VP64: T35S (5162 bp)

Table 5.7: List of Addgene plasmids used in this study

Name	Description
pICH86966 Addgene #48075	Used as the backbone of the shuttle vector; Contains the pVS1 and pMB1 replicons, KanR (bacterial) as well as a plant KanR under pNos.
pYTK001 Addgene #65108	Backbone for cloning all the parts listed in Table S1

Table 5.7 (continued)

pYTK002 Addgene #65109	Contains the connector LS; Used for the construction of TU1
pYTK003 Addgene #65110	Contains the connector L1; Used for the construction of TU2
pYTK004 Addgene #65111	Contains the connector L2; Used for the construction of TU3
pYTK005 Addgene #65112	Contains the connector L3; Used for the construction of TU4
pYTK006 Addgene #65113	Contains the connector L4; Used for the construction of TU5
pYTK007 Addgene #65114	Contains the connector L5; Used for the construction of TU6
pYTK067 Addgene #65174	Contains the connector R1; Used for the construction of TU1
pYTK068 Addgene #65175	Contains the connector R2; Used for the construction of TU2
pYTK069 Addgene #65176	Contains the connector R3; Used for the construction of TU3
pYTK070 Addgene #65177	Contains the connector R4; Used for the construction of TU4
pYTK071 Addgene #65178	Contains the connector R5; Used for the construction of TU5

Table 5.7 (continued)

pYTK072 Addgene #65179	Contains the connector RE; Used for the construction of TU6
pYTK095 Addgene #65202	Used as the backbone for the construction of any transcriptional unit.

Some concluding thoughts

As cells engineered with synthetic gene circuits find increasing utility in practical applications in the form of engineered probiotics and smart immunotherapy, moving forward there is a need for establishing a paradigm for the construction of such circuits with robust and predictable behavior across diverse organisms. In my opinion, the design of orthogonal control systems specially focused on transcriptional regulation will be essential in realizing this goal. Over the last few years, a number of strategies have been adopted to engineer such systems with the use of orthogonal transcription factors, novel sigma factors (prokaryotes), and occasionally using an entirely orthogonal RNA polymerase in the form of T7 RNA polymerase (T7 RNAP). After spending the last few years, trying to understand the nuances behind the circuit design incorporating these components, I have realized a few things that I feel can be of broad utility to the synthetic biology community.

First, while any arbitrary circuit topology can be designed on paper, construction of recombinant plasmids encoding them remains far from trivial. Majority of the circuits described in this work was only feasible because of the modular architecture that enabled their facile construction. The time and effort I spent in the design and characterization of such an architecture proved to be critical in realizing some of the seemingly impossible circuit designs. While a number of strategies have previously been described, I have attempted to unify most of the cloning efforts under a single framework (Chapters 4 and 5). These architectures will hopefully allow the reusability of genetic parts and plasmid

backbones which is essential as we attempt to characterize complex gene circuits across diverse species possibly under a universal paradigm.

The T7 RNAP based expression platform (homeostasis circuit) described in this work (Chapter 3 and 4) here is among the first steps in creating a universal gene expression platform, at least in prokaryotes. However, in its current form, it acts only as a constitutive expression platform. In a manner analogous to what was achieved in the early 2000s, I believe a number of circuit topologies need to be explored using T7 RNAP as the primary transcription engine. The most obvious next step will be the construction of inducible systems based on this platform that can be reliably ported across species. The significance of these systems might not be obvious at a first glance; however, we need to remember, that such gene switches were the cornerstone of most synthetic gene circuits that have been built in the last two decades. I imagine that in such a system, T7 RNAP promoters can be combined with transcription repressor operators to control the expression of a given output, while the expression of the respective transcriptional regulator is placed under the T7 RNAP control. The expression of both components can then be systematically optimized with the use of mutant T7 RNAP promoters similar to the strategy described in Chapter 4. Once a set of optimized expression parameters have been identified, they need to be validated across species. Such T7 RNAP based orthogonal gene switches can then be stitched together to construct more complex control systems similar to the ones that have traditionally relied on the host RNAP regulation. One of the interesting circuit topologies (and perhaps the simpler one) that can be built first will be the toggle switch. Given that the toggle switch architecture consists of two dueling repressors, it can be readily expanded

to work with T7 RNAP promoters. The characterization of these circuits will also require extensive modeling, not only in *E.coli* but also in other species. The portability of such circuits, and not just simple expression across diverse species will be a significant step in establishing genetic programs with 'universal' parts and modules.

Apart from these gene switches and modules, the expression of dCas9 based transcription factors and associated gRNA arrays, can also be placed under T7 RNAP control. Utilizing the programmability of the gRNA design, they can be readily adapted to modulate the expression of endogenous genes (apart from synthetic promoters), thus opening up the possibility of implementing large phenotypic changes under orthogonal transcription control. In particular, it has previously been shown in yeast that T7 RNAP expressed gRNAs can be used to regulate the activity of dCas9 based activators. Thus, in theory any control system based on T7 RNAP can be adapted to work in eukaryotes using the dCas9:VP64 to effectively transduce the signal. Overall, I can imagine large control systems based on T7 RNAP as the basic transcription engine, and the highly programmable effector protein such as dCas9 based transcription factors, enabling the construction of cross-kingdom circuity.

References

[1] Monod, J., and Jacob, F. (1961) Teleonomic mechanisms in cellular metabolism, growth, and differentiation, *Cold Spring Harb Symp Quant Biol 26*, 389-401.

[2] Ideker, T., Thorsson, V., Ranish, J. A., Christmas, R., Buhler, J., Eng, J. K., Bumgarner, R., Goodlett, D. R., Aebersold, R., and Hood, L. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network, *Science (New York, N.Y.) 292*, 929-934.

[3] Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabasi, A. L. (2000) The large-scale organization of metabolic networks, *Nature 407*, 651-654.

[4] Hartwell, L. H., Hopfield, J. J., Leibler, S., and Murray, A. W. (1999) From molecular to modular cell biology, *Nature 402*, C47-52.

[5] Alon, U. (2007) Network motifs: theory and experimental approaches, *Nat Rev Genet 8*, 450-461.

[6] Gossen, M., and Bujard, H. (1992) Tight control of gene expression in mammalian cells by tetracycline-responsive promoters, *Proceedings of the National Academy of Sciences of the United States of America 89*, 5547-5551.

[7] Figge, J., Wright, C., Collins, C. J., Roberts, T. M., and Livingston, D. M. (1988) Stringent regulation of stably integrated chloramphenicol acetyl transferase genes by E. coli lac repressor in monkey cells, *Cell 52*, 713-722.

[8] Hu, M. C., and Davidson, N. (1987) The inducible lac operator-repressor system is functional in mammalian cells, *Cell 48*, 555-566.

[9] Lutz, R., and Bujard, H. (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements, *Nucleic Acids Research 25*, 1203-1210.

[10] Tolle, F., Stucheli, P., and Fussenegger, M. (2019) Genetic circuitry for personalized human cell therapy, *Current opinion in biotechnology 59*, 31-38.

[11] Scheller, L., and Fussenegger, M. (2019) From synthetic biology to human therapy: engineered mammalian cells, *Current opinion in biotechnology 58*, 108-116.

[12] Sedlmayer, F., Aubel, D., and Fussenegger, M. (2018) Synthetic gene circuits for the detection, elimination and prevention of disease, *Nature biomedical engineering 2*, 399-415.

[13] Riglar, D. T., and Silver, P. A. (2018) Engineering bacteria for diagnostic and therapeutic applications, *Nature reviews. Microbiology 16*, 214-225.

[14] Kassaw, T. K., Donayre-Torres, A. J., Antunes, M. S., Morey, K. J., and Medford, J. I. (2018) Engineering synthetic regulatory circuits in plants, *Plant science : an international journal of experimental plant biology 273*, 13-22.

[15] Kiel, C., Yus, E., and Serrano, L. (2010) Engineering signal transduction pathways, *Cell 140*, 33-47.

[16] Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000) Construction of a genetic toggle switch in Escherichia coli, *Nature 403*, 339-342.

[17] Elowitz, M. B., and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators, *Nature 403*, 335-338.

[18] Kramer, B. P., and Fussenegger, M. (2005) Hysteresis in a synthetic mammalian gene network, *Proceedings of the National Academy of Sciences of the United States of America 102*, 9517-9522.
[19] Fung, E., Wong, W. W., Suen, J. K., Bulter, T., Lee, S. G., and Liao, J. C. (2005) A synthetic gene-metabolic oscillator, *Nature 435*, 118-122.
[20] Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H., and Weiss, R. (2005) A synthetic multicellular system for programmed pattern formation, *Nature 434*, 1130-1134.
[21] Basu, S., Mehreja, R., Thiberge, S., Chen, M. T., and Weiss, R. (2004) Spatiotemporal control of gene expression with pulse-generating networks, *Proceedings of the National Academy of Sciences of the United States of America 101*, 6355-6360.
[22] Stanton, B. C., Nielsen, A. A., Tamsir, A., Clancy, K., Peterson, T., and Voigt, C. A. (2014) Genomic mining of prokaryotic repressors for orthogonal logic gates, *Nature chemical biology 10*, 99-105.
[23] Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q. A., Tran, A. B., Paull, M., Keasling, J. D., Arkin, A. P., and Endy, D. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements, *Nature methods 10*, 354-360.
[24] Andrews, L. B., Nielsen, A. A. K., and Voigt, C. A. (2018) Cellular checkpoint control using programmable sequential logic, *Science (New York, N.Y.) 361*.
[25] Nielsen, A. A., Der, B. S., Shin, J., Vaidyanathan, P., Paralanov, V., Strychalski, E. A., Ross, D., Densmore, D., and Voigt, C. A. (2016) Genetic circuit design automation, *Science (New York, N.Y.) 352*, aac7341.
[26] Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C., and Voigt, C. A. (2012) Genetic programs constructed from layered logic gates in single cells, *Nature 491*, 249-253.
[27] Gander, M. W., Vrana, J. D., Voje, W. E., Carothers, J. M., and Klavins, E. (2017) Digital logic circuits in yeast with CRISPR-dCas9 NOR gates, *Nature communications 8*, 15459.
[28] Lee, M. E., DeLoache, W. C., Cervantes, B., and Dueber, J. E. (2015) A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly, *ACS synthetic biology 4*, 975-986.
[29] Kramer, B. P., Viretta, A. U., Daoud-El-Baba, M., Aubel, D., Weber, W., and Fussenegger, M. (2004) An engineered epigenetic transgene switch in mammalian cells, *Nature biotechnology 22*, 867-870.
[30] Chen, Y., Zhang, S., Young, E. M., Jones, T. S., Densmore, D., and Voigt, C. A. (2020) Genetic circuit design automation for yeast, *Nat Microbiol 5*, 1349-1360.
[31] Kim, H., Bojar, D., and Fussenegger, M. (2019) A CRISPR/Cas9-based central processing unit to program complex logic computation in human cells, *Proceedings of the National Academy of Sciences of the United States of America 116*, 7214-7219.
[32] Weber, W., and Fussenegger, M. (2009) Engineering of synthetic mammalian gene networks, *Chemistry & biology 16*, 287-297.

[33] Ellis, T., Wang, X., and Collins, J. J. (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions, *Nature biotechnology 27*, 465-471.
[34] Levskaya, A., Chevalier, A. A., Tabor, J. J., Simpson, Z. B., Lavery, L. A., Levy, M., Davidson, E. A., Scouras, A., Ellington, A. D., Marcotte, E. M., and Voigt, C. A. (2005) Synthetic biology: engineering Escherichia coli to see light, *Nature 438*, 441-442.
[35] Tabor, J. J., Salis, H., Simpson, Z. B., Chevalier, A. A., Levskaya, A., Marcotte, E. M., Voigt, C. A., and Ellington, A. D. (2009) A Synthetic Genetic Edge Detection Program, *Cell 137*, 1272-1281.
[36] Brophy, J. A., and Voigt, C. A. (2014) Principles of genetic circuit design, *Nature methods 11*, 508-520.
[37] Chappell, J., Watters, K. E., Takahashi, M. K., and Lucks, J. B. (2015) A renaissance in RNA synthetic biology: new mechanisms, applications and tools for the future, *Curr Opin Chem Biol 28*, 47-56.
[38] Pardee, K., Green, A. A., Ferrante, T., Cameron, D. E., DaleyKeyser, A., Yin, P., and Collins, J. J. (2014) Paper-based synthetic gene networks, *Cell 159*, 940-954.
[39] Green, A. A., Silver, P. A., Collins, J. J., and Yin, P. (2014) Toehold switches: de-novo-designed regulators of gene expression, *Cell 159*, 925-939.
[40] Mutalik, V. K., Qi, L., Guimaraes, J. C., Lucks, J. B., and Arkin, A. P. (2012) Rationally designed families of orthogonal RNA regulators of translation, *Nature chemical biology 8*, 447-454.
[41] Engstrom, M. D., and Pfleger, B. F. (2017) Transcription control engineering and applications in synthetic biology, *Synth Syst Biotechnol 2*, 176-191.
[42] Khalil, A. S., Lu, T. K., Bashor, C. J., Ramirez, C. L., Pyenson, N. C., Joung, J. K., and Collins, J. J. (2012) A synthetic biology framework for programming eukaryotic transcription functions, *Cell 150*, 647-658.
[43] Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., and Lim, W. A. (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression, *Cell 152*, 1173-1183.
[44] Li, Z., Zhang, D., Xiong, X., Yan, B., Xie, W., Sheen, J., and Li, J. F. (2017) A potent Cas9-derived gene activator for plant and mammalian cells, *Nature plants 3*, 930-936.
[45] Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., Thakore, P. I., Glass, K. A., Ousterout, D. G., Leong, K. W., Guilak, F., Crawford, G. E., Reddy, T. E., and Gersbach, C. A. (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors, *Nature methods 10*, 973-976.
[46] Nielsen, A. A. K., and Voigt, C. A. (2014) Multi-input CRISPR/Cas genetic circuits that interface host regulatory, *Molecular Systems Biology 10*.
[47] Liu, Q., Schumacher, J., Wan, X., Lou, C., and Wang, B. (2018) Orthogonality and Burdens of Heterologous AND Gate Gene Circuits in E. coli, *ACS synthetic biology 7*, 553-564.

[48] Cardinale, S., Joachimiak, M. P., and Arkin, A. P. (2013) Effects of genetic variation on the E. coli host-circuit interface, *Cell Rep 4*, 231-237.
[49] Kelly, J. R., Rubin, A. J., Davis, J. H., Ajo-Franklin, C. M., Cumbers, J., Czar, M. J., de Mora, K., Glieberman, A. L., Monie, D. D., and Endy, D. (2009) Measuring the activity of BioBrick promoters using an in vivo reference standard, *Journal of Biological Engineering 3*, 4.
[50] Cardinale, S., and Arkin, A. P. (2012) Contextualizing context for synthetic biology – identifying causes of failure of synthetic biological systems, *Biotechnology Journal 7*, 856-866.
[51] Cardinale, S., and Arkin, A. P. (2012) Contextualizing context for synthetic biology--identifying causes of failure of synthetic biological systems, *Biotechnology Journal 7*, 856-866.
[52] Studier, F. W., and Moffatt, B. A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes, *Journal of molecular biology 189*, 113-130.
[53] Conrad, B., Savchenko, R. S., Breves, R., and Hofemeister, J. (1996) A T7 promoter-specific, inducible protein expression system for Bacillus subtilis, *Molecular & general genetics : MGG 250*, 230-236.
[54] Brunschwig, E., and Darzins, A. (1992) A two-component T7 system for the overexpression of genes in Pseudomonas aeruginosa, *Gene 111*, 35-41.
[55] Ryu, M. H., Zhang, J., Toth, T., Khokhani, D., Geddes, B. A., Mus, F., Garcia-Costas, A., Peters, J. W., Poole, P. S., Ane, J. M., and Voigt, C. A. (2020) Control of nitrogen fixation in bacteria that associate with cereals, *Nat Microbiol 5*, 314-330.
[56] Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F., and Voigt, C. A. (2012) Modular control of multiple pathways using engineered orthogonal T7 polymerases, *Nucleic Acids Research 40*, 8773-8781.
[57] Imburgio, D., Rong, M., Ma, K., and McAllister, W. T. (2000) Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants, *Biochemistry 39*, 10419-10430.
[58] Rong, M., He, B., McAllister, W. T., and Durbin, R. K. (1998) Promoter specificity determinants of T7 RNA polymerase, *Proceedings of the National Academy of Sciences of the United States of America 95*, 515-519.
[59] Kochetkov, S. N., Rusakova, E. E., and Tunitskaya, V. L. (1998) Recent studies of T7 RNA polymerase mechanism, *FEBS letters 440*, 264-267.
[60] Ujvari, A., and Martin, C. T. (1996) Thermodynamic and kinetic measurements of promoter binding by T7 RNA polymerase, *Biochemistry 35*, 14574-14582.
[61] Meyer, A. J., Ellefson, J. W., and Ellington, A. D. (2014) Directed Evolution of a Panel of Orthogonal T7 RNA Polymerase Variants for in Vivo or in Vitro Synthetic Circuitry, *ACS synthetic biology*.
[62] Fernandez-Rodriguez, J., Moser, F., Song, M., and Voigt, C. A. (2017) Engineering RGB color vision into Escherichia coli, *Nature chemical biology 13*, 706-708.

[63] Segall-Shapiro, T. H., Meyer, A. J., Ellington, A. D., Sontag, E. D., and Voigt, C. A. (2014) A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase, *Molecular Systems Biology 10*, 742.
[64] Pu, J., Zinkus-Boltz, J., and Dickinson, B. C. (2017) Evolution of a split RNA polymerase as a versatile biosensor platform, *Nature chemical biology 13*, 432-438.
[65] Kushwaha, M., and Salis, H. M. (2015) A portable expression resource for engineering cross-species genetic circuits and pathways, *Nature communications 6*, 7832.
[66] Benton, B. M., Eng, W. K., Dunn, J. J., Studier, F. W., Sternglanz, R., and Fisher, P. A. (1990) Signal-mediated import of bacteriophage T7 RNA polymerase into the Saccharomyces cerevisiae nucleus and specific transcription of target genes, *Mol Cell Biol 10*, 353-360.
[67] Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., E, P. R. I., Lin, S., Kiani, S., Guzman, C. D., Wiegand, D. J., Ter-Ovanesyan, D., Braff, J. L., Davidsohn, N., Housden, B. E., Perrimon, N., Weiss, R., Aach, J., Collins, J. J., and Church, G. M. (2015) Highly efficient Cas9-mediated transcriptional programming, *Nature methods 12*, 326-328.
[68] Urlinger, S., Baron, U., Thellmann, M., Hasan, M. T., Bujard, H., and Hillen, W. (2000) Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity, *Proceedings of the National Academy of Sciences of the United States of America 97*, 7963-7968.
[69] Bae, K. H., Kwon, Y. D., Shin, H. C., Hwang, M. S., Ryu, E. H., Park, K. S., Yang, H. Y., Lee, D. K., Lee, Y., Park, J., Kwon, H. S., Kim, H. W., Yeh, B. I., Lee, H. W., Sohn, S. H., Yoon, J., Seol, W., and Kim, J. S. (2003) Human zinc fingers as building blocks in the construction of artificial transcription factors, *Nature biotechnology 21*, 275-280.
[70] Chavez, A., Tuttle, M., Pruitt, B. W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S. J., Cecchi, R. J., Kowal, E. J. K., Buchthal, J., Housden, B. E., Perrimon, N., Collins, J. J., and Church, G. (2016) Comparison of Cas9 activators in multiple species, *Nature methods 13*, 563-567.
[71] Farzadfard, F., Perli, S. D., and Lu, T. K. (2013) Tunable and multifunctional eukaryotic transcription factors based on CRISPR/Cas, *ACS synthetic biology 2*, 604-613.
[72] Morse, N. J., Wagner, J. M., Reed, K. B., Gopal, M. R., Lauffer, L. H., and Alper, H. S. (2018) T7 Polymerase Expression of Guide RNAs in vivo Allows Exportable CRISPR-Cas9 Editing in Multiple Yeast Hosts, *ACS synthetic biology 7*, 1075-1084.
[73] Lu, T. K., Khalil, A. S., and Collins, J. J. (2009) Next-generation synthetic gene networks, *Nature biotechnology 27*, 1139-1150.
[74] Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S., and Hasty, J. (2008) A fast, robust and tunable synthetic gene oscillator, *Nature 456*, 516-519.
[75] Tamsir, A., Tabor, J. J., and Voigt, C. A. (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires', *Nature 469*, 212-215.

[76] Regot, S., Macia, J., Conde, N., Furukawa, K., Kjellen, J., Peeters, T., Hohmann, S., de Nadal, E., Posas, F., and Sole, R. (2011) Distributed biological computation with multicellular engineered networks, *Nature 469*, 207-211.
[77] Purcell, O., and Lu, T. K. (2014) Synthetic analog and digital circuits for cellular computation and memory, *Current opinion in biotechnology 29*, 146-155.
[78] Daniel, R., Rubens, J. R., Sarpeshkar, R., and Lu, T. K. (2013) Synthetic analog computation in living cells, *Nature 497*, 619-623.
[79] Garamella, J., Marshall, R., Rustad, M., and Noireaux, V. (2016) The All E. coli TX-TL Toolbox 2.0: A Platform for Cell-Free Synthetic Biology, *ACS synthetic biology 5*, 344-355.
[80] Sun, Z. Z., Yeung, E., Hayes, C. A., Noireaux, V., and Murray, R. M. (2014) Linear DNA for rapid prototyping of synthetic biological circuits in an Escherichia coli based TX-TL cell-free system, *ACS synthetic biology 3*, 387-397.
[81] Zhang, C., Tsoi, R., and You, L. (2016) Addressing biological uncertainties in engineering gene circuits, *Integrative biology : quantitative biosciences from nano to macro 8*, 456-464.
[82] Zhang, D. Y., and Seelig, G. (2011) Dynamic DNA nanotechnology using strand-displacement reactions, *Nature chemistry 3*, 103-113.
[83] Qian, L., Winfree, E., and Bruck, J. (2011) Neural network computation with DNA strand displacement cascades, *Nature 475*, 368-372.
[84] Qian, L., and Winfree, E. (2011) Scaling up digital circuit computation with DNA strand displacement cascades, *Science (New York, N.Y.) 332*, 1196-1201.
[85] Seelig, G., Soloveichik, D., Zhang, D. Y., and Winfree, E. (2006) Enzyme-free nucleic acid logic circuits, *Science (New York, N.Y.) 314*, 1585-1588.
[86] Yurke, B., Turberfield, A. J., Mills, A. P., Jr., Simmel, F. C., and Neumann, J. L. (2000) A DNA-fuelled molecular machine made of DNA, *Nature 406*, 605-608.
[87] Chen, Y. J., Dalchau, N., Srinivas, N., Phillips, A., Cardelli, L., Soloveichik, D., and Seelig, G. (2013) Programmable chemical controllers made from DNA, *Nature nanotechnology 8*, 755-762.
[88] Zhang, D. Y., and Winfree, E. (2009) Control of DNA strand displacement kinetics using toehold exchange, *Journal of the American Chemical Society 131*, 17303-17314.
[89] Srinivas, N., Parkin, J., Seelig, G., Winfree, E., and Soloveichik, D. (2017) Enzyme-free nucleic acid dynamical systems, *Science (New York, N.Y.) 358*.
[90] Soloveichik, D., Seelig, G., and Winfree, E. (2010) DNA as a universal substrate for chemical kinetics, *Proceedings of the National Academy of Sciences of the United States of America 107*, 5393-5398.
[91] Kim, J., White, K. S., and Winfree, E. (2006) Construction of an in vitro bistable circuit from synthetic transcriptional switches, *Molecular Systems Biology 2*, 68.
[92] Subsoontorn, P., Kim, J., and Winfree, E. (2012) Ensemble Bayesian analysis of bistability in a synthetic transcriptional switch, *ACS synthetic biology 1*, 299-316.

[93] Padirac, A., Fujii, T., and Rondelez, Y. (2012) Bottom-up construction of in vitro switchable memories, *Proceedings of the National Academy of Sciences of the United States of America 109*, E3212-3220.
[94] Montagne, K., Plasson, R., Sakai, Y., Fujii, T., and Rondelez, Y. (2011) Programming an in vitro DNA oscillator using a molecular networking strategy, *Molecular Systems Biology 7*, 466.
[95] Kim, J., and Winfree, E. (2011) Synthetic in vitro transcriptional oscillators, *Molecular Systems Biology 7*, 465.
[96] Franco, E., Friedrichs, E., Kim, J., Jungmann, R., Murray, R., Winfree, E., and Simmel, F. C. (2011) Timing molecular motion and production with a synthetic transcriptional clock, *Proceedings of the National Academy of Sciences of the United States of America 108*, E784-793.
[97] Kim, J., Khetarpal, I., Sen, S., and Murray, R. M. (2014) Synthetic circuit for exact adaptation and fold-change detection, *Nucleic Acids Research 42*, 6078-6089.
[98] Weitz, M., Kim, J., Kapsner, K., Winfree, E., Franco, E., and Simmel, F. C. (2014) Diversity in the dynamical behaviour of a compartmentalized programmable biochemical oscillator, *Nature chemistry 6*, 295-302.
[99] Zadeh, J. N., Steenberg, C. D., Bois, J. S., Wolfe, B. R., Pierce, M. B., Khan, A. R., Dirks, R. M., and Pierce, N. A. (2011) NUPACK: Analysis and design of nucleic acid systems, *Journal of computational chemistry 32*, 170-173.
[100] Srinivas, N., Ouldridge, T. E., Šulc, P., Schaeffer, J. M., Yurke, B., Louis, A. A., Doye, J. P. K., and Winfree, E. (2013) On the biophysics and kinetics of toehold-mediated DNA strand displacement, *Nucleic Acids Research 41*, 10641-10658.
[101] Šulc, P., Ouldridge T , E., Romano, F., Doye J , P. K., and Louis A , A. (2015) Modelling Toehold-Mediated RNA Strand Displacement, *Biophysical Journal 108*, 1238-1247.
[102] Simpson, Z. B., Tsai, T. L., Nguyen, N., Chen, X., and Ellington, A. D. (2009) Modelling amorphous computations with transcription networks, *Journal of the Royal Society, Interface 6 Suppl 4*, S523-533.
[103] Babendure, J. R., Adams, S. R., and Tsien, R. Y. (2003) Aptamers switch on fluorescence of triphenylmethane dyes, *Journal of the American Chemical Society 125*, 14716-14717.
[104] Genot, A. J., Zhang, D. Y., Bath, J., and Turberfield, A. J. (2011) Remote toehold: a mechanism for flexible control of DNA hybridization kinetics, *Journal of the American Chemical Society 133*, 2177-2182.
[105] Green, A. A., Kim, J., Ma, D., Silver, P. A., Collins, J. J., and Yin, P. (2017) Complex cellular logic computation using ribocomputing devices, *Nature 548*, 117-121.
[106] Siegal-Gaskins, D., Tuza, Z. A., Kim, J., Noireaux, V., and Murray, R. M. (2014) Gene circuit performance characterization and resource usage in a cell-free "breadboard", *ACS synthetic biology 3*, 416-425.
[107] Pardee, K., Green, A. A., Takahashi, M. K., Braff, D., Lambert, G., Lee, J. W., Ferrante, T., Ma, D., Donghia, N., Fan, M., Daringer, N. M., Bosch, I., Dudley, D. M., O'Connor, D. H., Gehrke, L., and Collins, J. J. (2016) Rapid, Low-Cost

Detection of Zika Virus Using Programmable Biomolecular Components, *Cell 165*, 1255-1266.
[108] Pardee, K., Slomovic, S., Nguyen, P. Q., Lee, J. W., Donghia, N., Burrill, D., Ferrante, T., McSorley, F. R., Furuta, Y., Vernet, A., Lewandowski, M., Boddy, C. N., Joshi, N. S., and Collins, J. J. (2016) Portable, On-Demand Biomolecular Manufacturing, *Cell 167*, 248-259.e212.
[109] Dudley, Q. M., Karim, A. S., and Jewett, M. C. (2015) Cell-free metabolic engineering: biomanufacturing beyond the cell, *Biotechnology Journal 10*, 69-82.
[110] Carlson, E. D., Gan, R., Hodgman, C. E., and Jewett, M. C. (2012) Cell-free protein synthesis: applications come of age, *Biotechnology advances 30*, 1185-1194.
[111] Zawada, J. F., Yin, G., Steiner, A. R., Yang, J., Naresh, A., Roy, S. M., Gold, D. S., Heinsohn, H. G., and Murray, C. J. (2011) Microscale to Manufacturing Scale-up of Cell-Free Cytokine Production—A New Approach for Shortening Protein Production Development Timelines, *Biotechnology and bioengineering 108*, 1570-1578.
[112] Didovyk, A., Tonooka, T., Tsimring, L., and Hasty, J. (2017) Rapid and Scalable Preparation of Bacterial Lysates for Cell-Free Gene Expression, *ACS synthetic biology 6*, 2198-2208.
[113] Chen, Y. Y. (2011) From DNA to Targeted Therapeutics: Bringing Synthetic Biology to the Clinic, *3*, 106ps142.
[114] Ellefson, J. W., Meyer, A. J., Hughes, R. A., Cannon, J. R., Brodbelt, J. S., and Ellington, A. D. (2014) Directed evolution of genetic parts and circuits by compartmentalized partnered replication, *Nature biotechnology 32*, 97-101.
[115] Studier, F. W. (1991) Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system, *Journal of molecular biology 219*, 37-44.
[116] Chamberlin, M., McGrath, J., and Waskell, L. (1970) New RNA polymerase from Escherichia coli infected with bacteriophage T7, *Nature 228*, 227-231.
[117] McBride, K. E., Schaaf, D. J., Daley, M., and Stalker, D. M. (1994) Controlled expression of plastid transgenes in plants based on a nuclear DNA-encoded and plastid-targeted T7 RNA polymerase, *Proceedings of the National Academy of Sciences of the United States of America 91*, 7301-7305.
[118] Miroux, B., and Walker, J. E. (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels, *Journal of molecular biology 260*, 289-298.
[119] Iost, I., Guillerez, J., and Dreyfus, M. (1992) Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo, *Journal of bacteriology 174*, 619-622.
[120] Skinner, G. M., Baumann, C. G., Quinn, D. M., Molloy, J. E., and Hoggett, J. G. (2004) Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase. A single-molecule view of the transcription cycle, *The Journal of biological chemistry 279*, 3239-3244.
[121] Cheetham, G. M., Jeruzalmi, D., and Steitz, T. A. (1999) Structural basis for initiation of transcription from an RNA polymerase-promoter complex, *Nature 399*, 80-83.

[122] Chan, C. T., Lee, J. W., Cameron, D. E., Bashor, C. J., and Collins, J. J. (2016) 'Deadman' and 'Passcode' microbial kill switches for bacterial containment, *Nature chemical biology 12*, 82-86.
[123] Voigt, C. A. (2006) Genetic parts to program bacteria, *Current opinion in biotechnology 17*, 548-557.
[124] Parts, R. o. S. B.
[125] Cambray, G., Guimaraes, J. C., Mutalik, V. K., Lam, C., Mai, Q. A., Thimmaiah, T., Carothers, J. M., Arkin, A. P., and Endy, D. (2013) Measurement and modeling of intrinsic transcription terminators, *Nucleic Acids Research 41*, 5139-5148.
[126] Sleight, S. C., Bartley, B. A., Lieviant, J. A., and Sauro, H. M. (2010) Designing and engineering evolutionary robust genetic circuits, *Journal of Biological Engineering 4*, 12.
[127] Dubendorff, J. W., and Studier, F. W. (1991) Creation of a T7 autogene. Cloning and expression of the gene for bacteriophage T7 RNA polymerase under control of its cognate promoter, *Journal of molecular biology 219*, 61-68.
[128] Chen, H., Shiroguchi, K., Ge, H., and Xie, X. S. (2015) Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli, *Molecular Systems Biology 11*, 781.
[129] Hintsche, M., and Klumpp, S. (2013) Dilution and the theoretical description of growth-rate dependent gene expression, *Journal of Biological Engineering 7*, 22.
[130] Nath, K., and Koch, A. L. (1970) Protein degradation in Escherichia coli. I. Measurement of rapidly and slowly decaying components, *The Journal of biological chemistry 245*, 2889-2900.
[131] Gibson, D. G. (2011) Enzymatic assembly of overlapping DNA fragments, *Methods in enzymology 498*, 349-361.
[132] Medford, J. I., and Prasad, A. (2016) Towards programmable plant genetic circuits, *The Plant journal : for cell and molecular biology 87*, 139-148.
[133] Friedland, A. E., Lu, T. K., Wang, X., Shi, D., Church, G., and Collins, J. J. (2009) Synthetic gene networks that count, *Science (New York, N.Y.) 324*, 1199-1202.
[134] Mimee, M., Tucker, A. C., Voigt, C. A., and Lu, T. K. (2016) Programming a Human Commensal Bacterium, Bacteroides thetaiotaomicron, to Sense and Respond to Stimuli in the Murine Gut Microbiota, *Cell Syst 2*, 214.
[135] Costello, A., and Badran, A. H. (2021) Synthetic Biological Circuits within an Orthogonal Central Dogma, *Trends in biotechnology 39*, 59-71.
[136] Liu, C. C., Jewett, M. C., Chin, J. W., and Voigt, C. A. (2018) Toward an orthogonal central dogma, *Nature chemical biology 14*, 103-106.
[137] Kar, S., and Ellington, A. D. (2018) Construction of synthetic T7 RNA polymerase expression systems, *Methods (San Diego, Calif.) 143*, 110-120.
[138] Fonseca, J. P., Bonny, A. R., Kumar, G. R., Ng, A. H., Town, J., Wu, Q. C., Aslankoohi, E., Chen, S. Y., Dods, G., Harrigan, P., Osimiri, L. C., Kistler, A. L., and El-Samad, H. (2019) A Toolkit for Rapid Modular Construction of Biological Circuits in Mammalian Cells, *ACS synthetic biology 8*, 2593-2606.

[139] Engler, C., Youles, M., Gruetzner, R., Ehnert, T. M., Werner, S., Jones, J. D., Patron, N. J., and Marillonnet, S. (2014) A golden gate modular cloning toolbox for plants, *ACS synthetic biology 3*, 839-843.
[140] Segall-Shapiro, T. H., Sontag, E. D., and Voigt, C. A. (2018) Engineered promoters enable constant gene expression at any copy number in bacteria, *Nature biotechnology 36*, 352-358.
[141] Leonard, S. P., Perutka, J., Powell, J. E., Geng, P., Richhart, D. D., Byrom, M., Kar, S., Davies, B. W., Ellington, A. D., Moran, N. A., and Barrick, J. E. (2018) Genetic Engineering of Bee Gut Microbiome Bacteria with a Toolkit for Modular Assembly of Broad-Host-Range Plasmids, *ACS synthetic biology 7*, 1279-1290.
[142] Ou, B., Yang, Y., Tham, W. L., Chen, L., Guo, J., and Zhu, G. (2016) Genetic engineering of probiotic Escherichia coli Nissle 1917 for clinical application, *Applied microbiology and biotechnology 100*, 8693-8699.
[143] Loeschcke, A., and Thies, S. (2015) Pseudomonas putida-a versatile host for the production of natural products, *Applied microbiology and biotechnology 99*, 6197-6214.
[144] Din, M. O., Danino, T., Prindle, A., Skalak, M., Selimkhanov, J., Allen, K., Julio, E., Atolia, E., Tsimring, L. S., Bhatia, S. N., and Hasty, J. (2016) Synchronized cycles of bacterial lysis for in vivo delivery, *Nature 536*, 81-85.
[145] de Lange, O., Klavins, E., and Nemhauser, J. (2018) Synthetic genetic circuits in crop plants, *Current opinion in biotechnology 49*, 16-22.
[146] Weinberg, B. H., Pham, N. T. H., Caraballo, L. D., Lozanoski, T., Engel, A., Bhatia, S., and Wong, W. W. (2017) Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells, *Nature biotechnology 35*, 453-462.
[147] Belcher, M. S., Vuu, K. M., Zhou, A., Mansoori, N., Agosto Ramos, A., Thompson, M. G., Scheller, H. V., Loqué, D., and Shih, P. M. (2020) Design of orthogonal regulatory systems for modulating gene expression in plants, *Nature chemical biology 16*, 857-865.
[148] Schaumberg, K. A., Antunes, M. S., Kassaw, T. K., Xu, W., Zalewski, C. S., Medford, J. I., and Prasad, A. (2016) Quantitative characterization of genetic parts and circuits for plant synthetic biology, *Nature methods 13*, 94-100.
[149] Brückner, K., Schäfer, P., Weber, E., Grützner, R., Marillonnet, S., and Tissier, A. (2015) A library of synthetic transcription activator-like effector-activated promoters for coordinated orthogonal gene expression in plants, *The Plant journal : for cell and molecular biology 82*, 707-716.
[150] Weber, E., Engler, C., Gruetzner, R., Werner, S., and Marillonnet, S. (2011) A modular cloning system for standardized assembly of multigene constructs, *PloS one 6*, e16765.
[151] Banta, L. M., and Montenegro, M. (2008) Agrobacterium and Plant Biotechnology, In *Agrobacterium: From Biology to Biotechnology* (Tzfira, T., and Citovsky, V., Eds.), pp 73-147, Springer New York, New York, NY.

[152] Berg, R. H., and Beachy, R. N. (2008) Fluorescent protein applications in plants, *Methods Cell Biol 85*, 153-177.
[153] Benfey, P. N., and Chua, N. H. (1990) The Cauliflower Mosaic Virus 35S Promoter: Combinatorial Regulation of Transcription in Plants, *Science (New York, N.Y.) 250*, 959-966.
[154] Odell, J. T., Nagy, F., and Chua, N. H. (1985) Identification of DNA sequences required for activity of the cauliflower mosaic virus 35S promoter, *Nature 313*, 810-812.
[155] Kerppola, T. K. (2008) Bimolecular Fluorescence Complementation (BiFC) Analysis as a Probe of Protein Interactions in Living Cells, *Annual Review of Biophysics 37*, 465-487.
[156] Tindall, A. J., Waller, J., Greenwood, M., Gould, P. D., Hartwell, J., and Hall, A. (2015) A comparison of high-throughput techniques for assaying circadian rhythms in plants, *Plant Methods 11*, 32.
[157] Xiong, T. C., Sanchez, F., Briat, J.-F., Gaymard, F., and Dubos, C. (2016) Spatio-Temporal Imaging of Promoter Activity in Intact Plant Tissues, In *Plant Synthetic Promoters: Methods and Protocols* (Hehl, R., Ed.), pp 103-110, Springer New York, New York, NY.
[158] Xu, G., Greene, G. H., Yoo, H., Liu, L., Marqués, J., Motley, J., and Dong, X. (2017) Global translational reprogramming is a fundamental layer of immune regulation in plants, *Nature 545*, 487-490.
[159] Zhou, M., Wang, W., Karapetyan, S., Mwimba, M., Marqués, J., Buchler, N. E., and Dong, X. (2015) Redox rhythm reinforces the circadian clock to gate immune response, *Nature 523*, 472.
[160] Vaillant, I., Schubert, I., Tourmente, S., and Mathieu, O. (2006) MOM1 mediates DNA-methylation-independent silencing of repetitive sequences in Arabidopsis, *EMBO Rep 7*, 1273-1278.
[161] Matzke, M. A., Mette, M. F., and Matzke, A. J. (2000) Transgene silencing by the host genome defense: implications for the evolution of epigenetic control mechanisms in plants and vertebrates, *Plant Mol Biol 43*, 401-415.
[162] Ni, M., Cui, D., Einstein, J., Narasimhulu, S., Vergara, C. E., and Gelvin, S. B. (1995) Strength and tissue specificity of chimeric promoters derived from the octopine and mannopine synthase genes, *The Plant Journal 7*, 661-676.
[163] Matzke, M. A., Primig, M., Trnovsky, J., and Matzke, A. J. (1989) Reversible methylation and inactivation of marker genes in sequentially transformed tobacco plants, *Embo j 8*, 643-649.
[164] Cardinale, S., and Arkin, A. P. (2012) Contextualizing context for synthetic biology – identifying causes of failure of synthetic biological systems, *Biotechnology Journal 7*, 856-866.
[165] Kang, H.-G., Fang, Y., and Singh, K. B. (1999) A glucocorticoid-inducible transcription system causes severe growth defects in Arabidopsis and induces defense-related genes, *The Plant Journal 20*, 127-133.

[166] Amirsadeghi, S., McDonald, A. E., and Vanlerberghe, G. C. (2007) A glucocorticoid-inducible gene expression system can cause growth defects in tobacco, *Planta 226*, 453-463.
[167] Lowder, L. G., Paul, J. W., 3rd, and Qi, Y. (2017) Multiplexed Transcriptional Activation or Repression in Plants Using CRISPR-dCas9-Based Systems, *Methods in molecular biology (Clifton, N.J.) 1629*, 167-184.
[168] Bayer, T. S., and Smolke, C. D. (2005) Programmable ligand-controlled riboregulators of eukaryotic gene expression, *Nature Biotechnology 23*, 337-343.
[169] Kotula, J. W., Kerns, S. J., Shaket, L. A., Siraj, L., Collins, J. J., Way, J. C., and Silver, P. A. (2014) Programmable bacteria detect and record an environmental signal in the mammalian gut, *Proceedings of the National Academy of Sciences 111*, 4838.
[170] Jacobs, J. Z., Ciccaglione, K. M., Tournier, V., and Zaratiegui, M. (2014) Implementation of the CRISPR-Cas9 system in fission yeast, *Nature Communications 5*, 5344.
[171] Cruz, A. B., Bianchetti, R. E., Alves, F. R. R., Purgatto, E., Peres, L. E. P., Rossi, M., and Freschi, L. (2018) Light, Ethylene and Auxin Signaling Interaction Regulates Carotenoid Biosynthesis During Tomato Fruit Ripening, *Frontiers in Plant Science 9*.
[172] Lelli, K. M., Slattery, M., and Mann, R. S. (2012) Disentangling the many layers of eukaryotic transcriptional regulation, *Annual review of genetics 46*, 43-68.
[173] Ellwood, K., Huang, W., Johnson, R., and Carey, M. (1999) Multiple layers of cooperativity regulate enhanceosome-responsive RNA polymerase II transcription complex assembly, *Mol Cell Biol 19*, 2613-2623.
[174] Chen, L. (1999) Combinatorial gene regulation by eukaryotic transcription factors, *Curr Opin Struct Biol 9*, 48-55.
[175] Bellí, G., Garí, E., Piedrafita, L., Aldea, M., and Herrero, E. (1998) An activator/repressor dual system allows tight tetracycline-regulated gene expression in budding yeast, *Nucleic Acids Research 26*, 942-947.
[176] Perez-Pinera, P., Ousterout, D. G., Brunger, J. M., Farin, A. M., Glass, K. A., Guilak, F., Crawford, G. E., Hartemink, A. J., and Gersbach, C. A. (2013) Synergistic and tunable human gene activation by combinations of synthetic transcription factors, *Nature methods 10*, 239-242.
[177] Li, Y., Moore, R., Guinn, M., and Bleris, L. (2012) Transcription activator-like effector hybrids for conditional control and rewiring of chromosomal transgene expression, *Sci Rep 2*, 897.
[178] Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J. A., Somia, N. V., Bogdanove, A. J., and Voytas, D. F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting, *Nucleic Acids Research 39*, e82.
[179] Lowder, L. G., Zhang, D., Baltes, N. J., Paul, J. W., 3rd, Tang, X., Zheng, X., Voytas, D. F., Hsieh, T. F., Zhang, Y., and Qi, Y. (2015) A CRISPR/Cas9 Toolbox for

Multiplexed Plant Genome Editing and Transcriptional Regulation, *Plant physiology 169*, 971-985.
[180] Kocak, D. D., Josephs, E. A., Bhandarkar, V., Adkar, S. S., Kwon, J. B., and Gersbach, C. A. (2019) Increasing the specificity of CRISPR systems with engineered RNA secondary structures, *Nature biotechnology 37*, 657-666.
[181] Pandelakis, M., Delgado, E., and Ebrahimkhani, M. R. (2020) CRISPR-Based Synthetic Transcription Factors In Vivo: The Future of Therapeutic Cellular Programming, *Cell Syst 10*, 1-14.
[182] Yeo, N. C., Chavez, A., Lance-Byrne, A., Chan, Y., Menn, D., Milanova, D., Kuo, C. C., Guo, X., Sharma, S., Tung, A., Cecchi, R. J., Tuttle, M., Pradhan, S., Lim, E. T., Davidsohn, N., Ebrahimkhani, M. R., Collins, J. J., Lewis, N. E., Kiani, S., and Church, G. M. (2018) An enhanced CRISPR repressor for targeted mammalian gene regulation, *Nature methods 15*, 611-616.
[183] Kwon, D. Y., Zhao, Y. T., Lamonica, J. M., and Zhou, Z. (2017) Locus-specific histone deacetylation using a synthetic CRISPR-Cas9-based HDAC, *Nature communications 8*, 15315.
[184] Gonzalez, T. L., Liang, Y., Nguyen, B. N., Staskawicz, B. J., Loqué, D., and Hammond, M. C. (2015) Tight regulation of plant immune responses by combining promoter and suicide exon elements, *Nucleic Acids Research 43*, 7152-7161.
[185] Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W., and Chua, N.-H. (2006) Agrobacterium-mediated transformation of Arabidopsis thaliana using the floral dip method, *Nature Protocols 1*, 641-646.